



**TUGAS AKHIR - KS 141501**

# **SEGMENTASI PELANGGAN MENGGUNAKAN ANALISIS RFM DAN ALGORITMA FUZZY C-MEANS UNTUK MEMBANTU PENGELOLAAN HUBUNGAN PELANGGAN PADA PT. XYZ**

**AKBAR RACHMAD VIRGIAWAN  
NRP 5211 100 102**

**Dosen Pembimbing :**  
**Rully Agus Hendrawan, S.Kom., M.Eng.**  
**Retno Aulia Vinarti, S.Kom., M.Kom.**

**JURUSAN SISTEM INFORMASI  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember  
Surabaya 2015**



**TUGAS AKHIR - KS 141501**

# **CUSTOMER SEGMENTATION USING RFM ANALYSIS AND FUZZY-C-MEANS ALGORITHM TO HELP CUSTOMER RELATIONSHIP MANAGEMENT AT PT. XYZ**

**AKBAR RACHMAD VIRGIAWAN**  
NRP 5211 100 102

**SUPERVISOR:**  
Rully Agus Hendrawan, S.Kom., M.Eng.  
Retno Aulia Vinarti, S.Kom., M.Kom.

**JURUSAN SISTEM INFORMASI**  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember  
Surabaya 2015



## **LEMBAR PENGESAHAN**

### **SEGMENTASI PELANGGAN MENGGUNAKAN ANALISIS RFM DAN ALGORITMA FUZZY C- MEANS UNTUK MEMBANTU PENGELOLAAN HUBUNGAN PELANGGAN PADA PT. XYZ**

#### **TUGAS AKHIR**

Disusun Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada  
Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh:

**AKBAR RACHMAD VIRGIWAN**

**NRP. 5211 100 102**

Surabaya, 10 Juli 2016

**KETUA JURUSAN SISTEM INFORMASI**



**Dr. Eng. FEBRILIYAN SAMOPA, S.Kom., M.Kom.**

**NIP.19730219 199802 1 001**



## LEMBAR PERSETUJUAN

### SEGMENTASI PELANGGAN MENGGUNAKAN ANALISIS RFM DAN ALGORITMA FUZZY C- MEANS UNTUK MEMBANTU PENGELOLAAN HUBUNGAN PELANGGAN PADA PT. XYZ

#### TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada

Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh :

**AKBAR RACHMAD VIRGIAWAN**  
**NRP. 5211 100 102**

Disetujui Tim Penguji : Tanggal Ujian:

Periode Wisuda: September 2015.


**Rully Agus Hendrawan S.Kom., M.Eng.**

  
(Pembimbing I)

**Retno Aulia Vinarti, S.Kom., M.Kom.**

  
(Pembimbing II)

**Renny Pradina K., S.T., M.T.**

  
(Penguji I)

**Irmasari Hafidz, S.Kom., M.Sc.**

  
(Penguji II)

# **SEGMENTASI PELANGGAN MENGGUNAKAN ANALISIS RFM DAN ALGORITMA FUZZY C- MEANS UNTUK MEMBANTU PENGELOLAAN HUBUNGAN PELANGGAN PADA PT. XYZ**

**Nama Mahasiswa** : Akbar Rachmad Virgiawan  
**NRP** : 5211 100 102  
**Jurusan** : Sistem Informasi FTIf-ITS  
**Pembimbing 1** : Rully Agus Hendrawan S.Kom.,  
M.Eng.  
**Pembimbing 2** : Retno Aulia Vinarti, S.Kom., M.Kom.

## **ABSTRAK**

*Kesehatan adalah kebutuhan yang paling utama setiap makhluk hidup, jika badan tidak sehat maka kita tidak akan bisa beraktifitas seperti biasanya. Oleh karena itu, untuk menjaga kesehatan dan mengembalikan kesehatan seseorang harus membayar dengan mahal. Karena hal itu maka bisnis produsen obat menjadi bisnis yang menggiurkan dengan keuntungan yang didapat. Potensi ini ingin dimaksimalkan oleh PT. XYZ untuk memperoleh keuntungan yang besar.*

*Untuk memperoleh keuntungan yang besar itu salah satu cara adalah memahami pelanggan. Salah satu cara untuk memahami pelanggan adalah dengan melakukan segmentasi pelanggan dan mencari karakteristik pelanggan. Cara itu dapat membantu PT. XYZ untuk lebih mengerti seperti apa pelanggan.*

*Segmentasi pelanggan dapat dilakukan dengan beberapa cara. Pada tugas akhir ini, segmentasi pelanggan akan dilakukan dengan kombinasi analisis RFM pada data transaksi pelanggan, Fuzzy C-Means Clustering (FCM), dan validasi SSE untuk mendapatkan berapa segmen pelanggan. Setiap segmen pelanggan akan dicari karakteristik dari setiap segmen dengan menggunakan analisis geografis, demografis, dan behavior.*

*Hasil dari pengerjaan tugas akhir ini menunjukkan terdapat enam kelompok pelanggan dengan enam karakteristiknya yang berbeda yaitu pelanggan terbaik, pelanggan yang mempunyai nilai, pelanggan yang sering berbelanja, pelanggan baru, pelanggan tidak menentu, dan pelanggan hilang. Telah dijelaskan pula karakteristik hasil dari analisis geografis, demografis, dan behavior.*

***Kata Kunci: Segmentasi Pelanggan, Karakteristik Pelanggan, Analisis RFM, Fuzzy C-Means Clustering, validasi SSE, Analisis Demografi, Analisis Geografi, Analisis Behavior.***

# **CUSTOMER SEGMENTATION USING RFM ANALYSIS AND FUZZY-C-MEANS ALGORITHM TO HELP CUSTOMER RELATIONSHIP MANAGEMENT AT PT. XYZ**

**Student Name : Akbar Rachmad Virgiawan**

**NRP : 5211 100 102**

**Department : Information System FTIf-ITS**

**Supervisor 1 : Rully Agus Hendrawan S.Kom., M.Eng.**

**Supervisor 2 : Retno Aulia Vinarti, S.Kom., M.Kom.**

## **ABSTRACT**

*Health is the most important needs of every living creature, if the body is not healthy then we will not be able to indulge as usual. Therefore, to maintain and restore the health of a person's health should pay dearly. Because it was the manufacturer of the drug business into a lucrative business with profits. This potential is wanted maximized by PT. XYZ to gain huge profits.*

*To obtain huge profits, one way is to understand the customer. One way to understand the customer is to perform customer segmentation and look for the characteristics of the customer. The way it can help PT. XYZ to better understand what the customers.*

*Customer segmentation can be done in several ways. In this thesis, customer segmentation will be done with a combination of RFM analysis on the customer transaction data, Fuzzy C-Means Clustering (FCM), and validation SSE to get how many customer segments. Each customer segment will be sought characteristics of each segment using geographic analysis, demographic, and behavior.*

*Results of this final project shows there are six groups of customers with six different characteristics that best customers, customers that have value, customers who frequently shop, new customers, customer erratic and lost customers. Have also described the results of the analysis of*

*the characteristics of the geographic, demographic, and behavior.*

**Keywords:** *Customer Segmentation, Customer Characteristic, RFM analysis, Fuzzy C-Means Clustering, SSE validation, Demographic analysis, Geographic analysis, Behavior analysis.*



## KATA PENGANTAR

Alhamdulillahirobbil ‘alamiin. Allahumma sholli’alaa Muhammad, wa ‘alaa aali sayyidina Muhammad. Tiada Dzat yang mampu menolong selain Allah SWT sehingga penulis dapat menyelesaikan buku tugas akhir dengan judul:

### **SEGMENTASI PELANGGAN MENGGUNAKAN ANALISIS RFM DAN ALGORITMA FUZZY C-MEANS UNTUK MEMBANTU PENGELOLAAN HUBUNGAN PELANGGAN PADA PT. XYZ**

yang merupakan salah satu syarat kelulusan pada Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya.

Dalam pengerjaan tugas akhir yang berlangsung selama bulan Februari – Juli 2015 penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada yang senantiasa terlibat secara langsung memberikan bantuan dan dukungan dalam pengerjaan tugas akhir ini :

- 1) Allah SWT yang telah memberikan kesehatan, kemudahan dan kesempatan untuk bisa menyelesaikan tugas akhir ini.
- 2) Kepada PT XYZ cabang Semarang yang bersedia memberikan data penjualan mereka sebagai data untuk pengerjaan tugas akhir ini.
- 3) Kepada Bapak Febriliyan Samopa selaku ketua jurusan sistem informasi.
- 4) Kepada Bapak Rully A. Hendrawan dan Ibu Retno Aulia Vinarti, selaku dosen pembimbing. Terima kasih atas segala bimbingan dan waktu yang telah diberikan untuk membantu penulis dalam menyelesaikan tugas akhir ini.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penulisan tugas akhir ini, oleh karena itu penulis menerima saran dan kritik yang membangun. Namun penulis berharap bahwa tugas akhir ini dapat bermanfaat bagi semua pihak yang membutuhkan.

Surabaya, Juni 2015

Penulis

## DAFTAR ISI

ABSTRAK .....	v
ABSTRACT .....	vii
KATA PENGANTAR .....	ix
DAFTAR ISI .....	xi
DAFTAR GAMBAR .....	xiii
DAFTAR TABEL .....	xv
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang Masalah .....	1
1.2. Rumusan Masalah .....	2
1.3. Batasan Tugas Akhir .....	2
1.4. Tujuan Penelitian .....	3
1.5. Manfaat Penelitian .....	3
1.6. Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI .....	5
2.1. Penelitian Sebelumnya .....	5
2.2. Dasar Teori .....	8
2.2.1. Customer Relationship Management .....	8
2.2.2. Customer Profiling and Segmentation .....	9
2.2.3. Model RFM .....	12
2.2.4. Normalisasi min-max .....	13
2.2.5. Penggalan Data .....	14
2.2.6. Clustering .....	15
2.2.7. <i>Validasi</i> .....	18
2.2.8. Aplikasi R untuk Proses Clustering, Validasi, dan Visualisasi .....	20
BAB III METODOLOGI PENELITIAN .....	25
3.1. Studi Literatur .....	25
3.2. Pengumpulan Data dan <i>Preprocessing</i> Data .....	25
3.3. Proses <i>Clustering</i> dengan Algoritma Fuzzy C-Means .....	26
3.4. Validasi Hasil <i>Cluster</i> dengan SSE .....	28
3.5. Analisis Karakteristik Pelanggan .....	28
3.6. Penyusunan Buku Tugas Akhir .....	28
BAB IV PERANCANGAN .....	29
4.1. Pengumpulan Data .....	29
4.2. <i>Praprocessing</i> Data .....	32
4.3. Mencari Nilai Atribut RFM .....	34

4.3.1.	Mencari Nilai <i>Recency</i> ( <i>R</i> ).....	34
4.3.2.	Mencari Nilai <i>Frequency</i> ( <i>F</i> ) .....	35
4.3.3.	Mencari Nilai <i>Monetary</i> ( <i>M</i> ) .....	35
4.4.	Normalisasi RFM Menggunakan Metode min-max ....	36
4.4.1.	<i>Recency</i> .....	37
4.4.2.	<i>Frequency</i> .....	37
4.4.3.	<i>Monetary</i> .....	38
BAB V IMPLEMENTASI.....		39
5.1.	<i>Clustering</i> menggunakan Algoritma Fuzzy C-Means .	39
5.2.	Validasi dengan Metode SSE .....	43
5.3.	Validasi dengan DBI.....	45
5.4.	Pembuatan Visualisasi untuk Membantu Analisis.....	47
5.4.1.	Visualisasi Scatter Plot.....	47
5.4.2.	Visualisasi dengan <i>googleVis</i> .....	49
5.4.3.	Visualisasi dengan <i>HeatMap</i> .....	52
BAB VI HASIL DAN PEMBAHASAN .....		55
6.1.	Analisis Hasil <i>Clustering</i> .....	55
6.1.1.	Analisis Antar Atribut RFM.....	55
6.1.2.	Hasil Analisis Geografi .....	61
6.1.3.	Hasil Analisis Tipe Pelanggan .....	68
6.1.4.	Hasil Analisis <i>Behavior</i> .....	73
6.2.	Menemukan Karakteristik Pelanggan .....	84
BAB VII KESIMPULAN DAN SARAN .....		89
7.1.	Kesimpulan .....	89
7.2.	Saran .....	91
DAFTAR PUSTAKA .....		93
BIODATA PENULIS .....		97
LAMPIRAN A .....		A-1
LAMPIRAN B .....		B-1
LAMPIRAN C .....		C-1
Ucapan Terima Kasih.....		114



## DAFTAR TABEL

Tabel 2.1 Penelitian Sebelumnya .....	6
Tabel 2.2 Contoh Data RFM Pelanggan.....	12
Tabel 4.1 Atribut dan Tipe Data Mentah.....	29
Tabel 4.2 Cuplikan Data Mentah.....	31
Tabel 4.3 Hasil Penyeleksian Atribut .....	32
Tabel 4.4 Contoh Pembersihan Data .....	33
Tabel 4.5 Cuplikan Nilai RFM Pelanggan .....	36
Tabel 4.6 Cuplikan Nilai <i>Recency</i> Pelanggan yang Telah Dinormalisasi.....	37
Tabel 4.7 Cuplikan Nilai <i>Frquency</i> Pelanggan yang Telah Dinormalisasi.....	37
Tabel 4.8 Cuplikan Nilai <i>Monetary</i> Pelanggan yang Telah Dinormalisasi.....	38
Tabel 5.1 Cuplikan Nilai RFM yang Telah Dinormalisasi .....	39
Tabel 5.2 <i>Centroid</i> untuk C=3 .....	41
Tabel 5.3 <i>Centroid</i> untuk C=4.....	41
Tabel 5.4 <i>Centroid</i> untuk C=5.....	41
Tabel 5.5 <i>Centroid</i> untuk C=6.....	42
Tabel 5.6 <i>Centroid</i> untuk C=7.....	42
Tabel 5.7 <i>Centroid</i> untuk C=8.....	42
Tabel 5.8 Cuplikan Hasil Seleksi Atribut untuk Membuat Visualisasi Map .....	50
Tabel 5.9 Cuplikan Hasil Penambahan Atribut <i>Longitude</i> dan <i>Latitude</i> .....	51
Tabel 5.10 Cuplikan Data yang Digunakan untuk Membuat Visualisasi <i>HeatMap</i> .....	53
Tabel 6.1 analisis <i>cluster</i> antara atribut <i>recency&amp;frequency</i> .....	56
Tabel 6.2 analisis <i>cluster</i> antara atribut <i>recency&amp;monetary</i> .....	58
Tabel 6.3 analisis <i>cluster</i> antara atribut <i>frequency &amp; monetary</i> ..	60
Tabel 6.4 Kesimpulan Hasil Analisis Geografi .....	66
Tabel 6.5 Jumlah Tipe Pelanggan Apotik dan Rumah Sakit pada Setiap Klaster .....	70

Tabel 6.6 Jumlah Tipe Pelanggan Klinik dan Perusahaan/Yayasan pada Setiap Klaster .....	71
Tabel 6.7 Analisis Berdasarkan Kualitas Pelanggan (Pareto) .....	71
Tabel 6.8 Transaksi klaster 1 per minggu .....	74
Tabel 6.9 Skala transaksi klaster 1 per minggu .....	75
Tabel 6.10 Nilai Ekstrim pada Klaster 1 .....	76
Tabel 6.11 Transaksi klaster 2 per minggu .....	76
Tabel 6.12 Nilai Ekstrim pada Klaster 2 .....	76
Tabel 6.13 Skala transaksi klaster 2 per minggu .....	77
Tabel 6.14 Transaksi klaster 3 per minggu .....	78
Tabel 6.15 Skala transaksi klaster 3 per minggu .....	78
Tabel 6.16 Nilai Ekstrim pada Klaster 3 .....	79
Tabel 6.17 Transaksi klaster 4 per minggu .....	79
Tabel 6.18 Skala transaksi klaster 4 per minggu .....	80
Tabel 6.19 Nilai Ekstrim pada Klaster 4 .....	80
Tabel 6.20 Transaksi klaster 5 per minggu .....	81
Tabel 6.21 Skala transaksi klaster 5 per minggu .....	81
Tabel 6.22 Nilai Ekstrim pada Klaster 5 .....	82
Tabel 6.23 Transaksi klaster 6 per minggu .....	83
Tabel 6.24 Skala transaksi klaster 6 per minggu .....	83
Tabel 6.25 Nilai Ekstrim pada Klaster 6 .....	84
Tabel 6.26 Tabel karakteristik pelanggan .....	85

## DAFTAR GAMBAR

Gambar 2.1 Contoh Hasil Validasi SSE.....	22
Gambar 2.2 Contoh Map dengan SHP .....	23
Gambar 3.1 Metodologi Penelitian .....	27
Gambar 5.1 Hasil Validasi SSE pada Aplikasi R.....	44
Gambar 5.2 Hasil Validasi DBI pada Aplikasi R.....	47
Gambar 5.3 Langkah Pembuatan Visualisasi dengan <i>Package</i> GoogleVis .....	49
Gambar 6.1 Scatter plot 3D (RFM).....	56
Gambar 6.2 Lokasi Pelanggan Klaster 1 .....	62
Gambar 6.3 Lokasi Pelanggan Klaster 2 .....	63
Gambar 6.4 Lokasi Pelanggan Klaster 3 .....	64
Gambar 6.5 Lokasi Pelanggan Klaster 4 .....	65
Gambar 6.6 Lokasi Pelanggan Klaster 5 .....	65
Gambar 6.7 Lokasi Pelanggan Klaster 6 .....	66
Gambar 6.8 Hasil analisis demografi .....	69
Gambar 6.9 Urutan Tipe Pelanggan Berdasarkan Indeks RFM dari Terbaik Sampai Terburuk.....	72

# **BAB I**

## **PENDAHULUAN**

Pada bagian pendahuluan akan dibahas mengenai latar belakang, perumusan masalah, batasan, tujuan, dan manfaat Tugas Akhir.

### **1.1. Latar Belakang Masalah**

Pada tahun 2011, pasar farmasi nasional mencapai Rp 43,08 triliun. Pasar tersebut diperkirakan bisa tumbuh sekitar 13% menjadi Rp 48,61 triliun pada tahun 2012 [1]. Hal ini membuktikan bahwa kesehatan merupakan kebutuhan yang paling penting dari setiap manusia. Karena dengan tubuh yang sehat setiap waktu yang mereka punya akan dapat mereka berdayakan dengan baik, apalagi arti kesehatan bagi masyarakat di Negara maju, benar-benar waktu sudah menjadi uang, tuntutan ekonomi yang tinggi menyebabkan mereka enggan bermalasan dalam menjalani hidup, itu sebabnya banyak dari masyarakat Negara maju paham betul manfaat kesehatan bagi hidup mereka masing-masing [2].

Potensi pasar yang sangat besar ini membuat PT. XYZ ingin memanfaatkannya untuk mendapatkan profit yang sangat besar pula. PT. XYZ bergerak dibidang produksi obat-obatan yang sudah berskala nasional. PT. XYZ memiliki visi untuk menjadi perusahaan kesehatan yang paling baik di Indonesia maupun dunia, maka PT. XYZ yang telah memiliki pelanggan sebanyak 475 apotek yang ada di daerah Jawa Tengah harus meningkatkan pengelolaan hubungan pelanggan untuk menguasai pasar ini. Di Indonesia, PT. Kalbe Farma Tbk menempati posisi kedua dalam penguasaan market share. Pada peringkat pertama terdapat Sanbe. [1]

PT. XYZ selama ini belum menerapkan pengelolaan hubungan pelanggan yang baik. PT. XYZ juga belum memberikan promosi atau reward bagi pelanggan yang loyal. Tetapi dengan semakin beratnya persaingan untuk memperebutkan market share farmasi, mau tidak mau PT. XYZ harus melakukan hal itu.



Selama ini cara menyelesaikan masalah di atas adalah dengan metode penggalan data transaksi pembelian oleh pelanggan. Metode ini dapat mengelompokkan pelanggan menjadi beberapa segmen yang dibedakan dengan potensi yang ada pada setiap pelanggan. Sebelum segmen dilakukan, perhitungan potensi pelanggan dilakukan terlebih dahulu dengan menggunakan model RFM (Recency, Frequency, Monetary). [3]

Segmentasi pelanggan biasanya dilakukan dengan algoritma K-Means, namun pada tugas akhir kali ini akan digunakan algoritma Fuzzy C-Means (FCM). Penggunaan algoritma FCM pada tugas akhir kali ini dilakukan karena algoritma FCM memiliki nilai entropy lebih kecil daripada algoritma K-Means. Sedangkan nilai purity dari algoritma FCM lebih baik daripada algoritma K-Means. [4]

## **1.2. Rumusan Masalah**

Rumusan masalah yang diangkat dalam tugas akhir ini adalah:

1. Terbagi ke dalam berapa segmenkah pelanggan perusahaan PT. XYZ?
2. Bagaimana karakteristik tiap segmen pelanggan perusahaan PT. XYZ?
3. Segmen berapakah yang mempunyai nilai RFM paling baik dan paling buruk?
4. Tipe pelanggan apa yang harus diperhatikan lebih oleh PT. XYZ?

## **1.3. Batasan Tugas Akhir**

Batasan dalam pengerjaan tugas akhir ini adalah:

1. Sumber data yang digunakan adalah data transaksi harian yang ada pada PT. XYZ selama 23 bulan yaitu mulai dari bulan Januari 2013 sampai bulan November 2014.
2. Studi kasus PT. XYZ yang berada pada kota Semarang Provinsi Jawa Tengah. Cabang Semarang mencakup: Kota Semarang, Kota Kendal, Kota Ungaran, dan

beberapa kota lainnya yang berada di sekitar kota Semarang.

#### **1.4. Tujuan Penelitian**

Tujuan dari tugas akhir ini adalah mengetahui karakteristik segmen pelanggan dengan melakukan pengelompokan menggunakan Fuzzy C-Means berdasarkan data transaksi belanja pelanggan yang dianalisis menggunakan model RFM, analisis geografi, analisis tipe pelanggan, dan analisis *behavior*.

#### **1.5. Manfaat Penelitian**

Tugas akhir ini diharapkan dapat memberikan manfaat untuk perusahaan PT. XYZ dalam memberikan gambaran segmentasi pelanggan sebagai masukan bagi perusahaan dalam menyusun dalam menyusun strategi hubungan pelanggan sehingga dapat memicu peningkatan profit.

#### **1.6. Sistematika Penulisan**

Dalam tugas akhir ini, sistematika penulisan laporan disesuaikan dengan pelaksanaan penelitian dan saling berhubungan antara satu dengan yang lainnya. Penulisan ini dibagi menjadi tujuh bab dan masing-masing bab terdiri dari beberapa sub bab untuk memberikan penjelasan yang lebih rinci. Sistematika penulisan laporan penelitian tugas akhir ini dijelaskan sebagai berikut:

### **BAB I PENDAHULUAN**

Pada bab ini dijelaskan mengenai latar belakang, perumusan masalah, batasan masalah tugas akhir, manfaat tugas akhir, tujuan tugas akhir dan sistematika penulisan yang diterapkan dalam memaparkan tugas akhir.

### **BAB II TINJAUAN PUSTAKA**

Pada bab ini akan dijelaskan tentang penelitian sebelumnya dan referensi-referensi yang berkaitan dengan tugas akhir, antara lain uraian mengenai *Customer Relationship Management*, *Customer Segmentation and Profiling*, Model RFM, normalisasi data, penggalan data, *clustering*, algoritma Fuzzy C-Means, SSE, dan aplikasi R.

### **BAB III METODE PENELITIAN**

Bab ini berisi penjelasan mengenai masing-masing tahap dalam pembuatan tugas akhir, mulai dari studi literatur, pengumpulan dan *preprocessing* data, proses clustering dengan menggunakan algoritma Fuzzy C-Means, validasi menggunakan SSE, analisis karakteristik pelanggan, hingga penyusunan tugas akhir.

### **BAB IV PERANCANGAN**

Bab ini dijelaskan mengenai penyelesaian tahap pengumpulan serta penyiapan data sebelum lebih lanjut diolah menggunakan Fuzzy C-Means *clustering*.

### **BAB IV IMPLEMENTASI**

Pada bab ini dijelaskan mengenai proses *clustering* dengan menggunakan algoritma Fuzzy C-Means clustering dan divalidasi dengan menggunakan SSE untuk menentukan jumlah segmen pelanggan yang maksimal.

### **BAB VI HASIL DAN PEMBAHASAN**

Bab ini berisi analisis hasil clustering dengan cara analisis geografis, analisis demografi, analisis behavior, dan analisis antar atribut RFM. Pada bab ini juga akan ditampilkan karakteristik pelanggan dari hasil analisis yang telah dilakukan.

### **BAB VII KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dan saran dari seluruh pengerjaan yang telah dilakukan untuk dibandingkan dengan tujuan dan permasalahan yang sudah dibuat pada bab pendahuluan.



## **BAB II**

### **TINJAUAN PUSTAKA DAN DASAR TEORI**

Pada bagian ini akan dibahas mengenai hasil studi penelitian sebelumnya dan dan teori-teori yang mendukung dalam pengerjaan tugas akhir.

#### **2.1. Penelitian Sebelumnya**

Pada sub bab tinjauan pustaka ini berisi tentang studi-studi atau penelitian-penelitian yang pernah dilakukan dan berkaitan dengan tugas akhir ini. Pada tinjabuan pustaka ini terdapat empat penelitian sebelumnya yang membahas tentang pembuatan segmentasi pelanggan menggunakan teknik penggalian data, model RFM, penggunaan algoritma Fuzzy C-Means, dan visualisasi dari analisis RFMnya.

Untuk studi literature yang pertama, penulis menggunakan data transaksi harian yang ada di PT. XYZ. Dari data tersebut dilakukan pembersihan dan normalisasi data. Setelah itu penulis melakukan model RFM yang diberi weighted dengan metode AHP. Setelah itu menentukan C klasternya dengan menggunakan algoritma fanny. Lalu mengklaster data tersebut dengan menggunakan Fuzzy C-Means dan menentukan peringkat nilai hidup pelanggan dengan menggunakan metode TOPSIS. Sedangkan pada tinjauan pustaka kedua tahapannya hampir sama pada tinjauan pustaka yang pertama. Namun bedanya model RFMnya biasa dan metode klaster yang dilakukan adalah K-Means.

Pada tinjauan pustaka yang ketiga yaitu membandingkan K-Means dan FCM dengan metode pengukuran validasi eksternal yaitu dengan purity dan entropy. Lalu pada tinjauan pustaka keempat, prosesnya sama dengan tinjauan pustaka satu dan dua. Namun perbedaannya ada pada penggunaan metode Ward untuk menentukan k dan visualisasi menggunakan Tableau.

**Tabel 2.1 Penelitian Sebelumnya**

No	Judul	Penulis	Data	Metode	Keluaran
1	EVALUASI DAN REKOMENDASI PENINGKATAN NILAI HIDUP PELANGGAN MENGGUNAKAN MODEL RFM DAN TEKNIK PENGALIAN DATA: STUDI KASUS PT. XYZ	Dyah Puspita Dewi	Data Pelanggan yang melakukan transaksi dengan Produk Printer	Weighted RFM Algoritma Fanny Fuzzy C-Means  TOPSIS	Rekomendasi strategi pemasaran dan rekomendasi produk
2	DATA MINING USING RFM ANALYSIS	Derya Birant	Data transaksi pada e- commerce dan Data Pelanggan	Model RFM K-Means Classification Association Rule	Segmen Pelanggan dan rekomendasi produk
3	COMPARISON OF PURITY AND ENTROPY OF K-MEANS AND FUZZY C-MEANS CLUSTERING	Satya Chaitanya Sripada	Data penyakit Hepatitis, Diabetes, Penyakit Hati, Penyakit Jantung, Tumor Otak, Penyakit Mata	Pengukuran validasi Eksternal	Fuzzy C-Means memiliki nilai entropi dan purity yang lebih baik

No	Judul	Penulis	Data	Metode	Keluaran
4	VISUALISASI SEGMENTASI FITUR ATM MENGGUNAKAN ANALISIS RFM DAN ALGORITMA K-MEANS UNTUK MENDUKUNG EVALUASI FITUR ATM DI PT. BANK SYARIAH	Eka Ayu Kurniati	Data transaksi pada tiap fitur ATM	Weighted RFM Algoritma K-Means Indeks Dunn Indeks R-Squared SSE Ward Visualisasi	Visualisasi model RFM dengan scatter plot lalu mendefinisikan tipe tiap segmen

## 2.2. Dasar Teori

Pada sub bab dasar teori akan dijelaskan tentang teori-teori mengenai Pengelolaan Hubungan Pelanggan (PHP), Model RFM, Penggalan data, *Clustering*, *Sum of Square Error*, Visualisasi, dan Aplikasi Tableau.

### 2.2.1. Customer Relationship Management

CRM adalah sebuah pendekatan strategis yang berkonsentrasi untuk meningkatkan value terhadap shareholder melalui penanganan hubungan yang tepat dengan customer utama dan juga dengan kelompok pelanggan lainnya. CRM pertama kali populer di tahun 1990, pada tahun itu CRM sering digunakan untuk mendeskripsikan solusi pelanggan berbasis teknologi. CRM memberikan kesempatan yang lebih baik untuk menggunakan data dan informasi dalam rangka untuk mengerti tipe pelanggan dan membuat nilai yang lebih terhadap pelanggan. [5]

Pada CRM terdapat dua strategi yang sangat berkaitan erat, yaitu strategi bisnis dan strategi pelanggan. Pada tugas akhir kali ini fokus pada strategi pelanggan itu sendiri. Strategi pelanggan memiliki proses untuk memeriksa basis pelanggan yang sudah ada dan pelanggan yang potensial, tidak hanya memeriksa tetapi juga harus mengidentifikasi bentuk segmentasi pelanggan yang paling tepat. Perusahaan harus mengidentifikasinya untuk membantu perusahaan dalam mengambil tindakan dalam melayani para pelanggan. [5]

Segmentasi pelanggan adalah pengelompokan atau klasifikasi pelanggan yang potensial dengan pelanggan yang saat ini ada. Pengelompokan atau klasifikasi dilakukan dengan melihat reaksi pasar pelanggan atau transaksi yang dilakukan oleh pelanggan. [6] Segmentasi pelanggan ini bertujuan untuk menyesuaikan potensi dari pelanggan tersebut dengan layanan serta strategi pemasaran yang dilakukan agar strategi pemasaran dan layanan yang diberikan tepat guna. [7]

### 2.2.2. Customer Profiling and Segmentation

Banyak perusahaan yang berusaha bertahun-tahun untuk memompa banyak produksi yang mereka agar bisa terjual di pasar tanpa petunjuk seperti apa pelanggan terbaik mereka. Karena itu *Customer Profiling* dan *Segmentation* adalah langkah pertama yang baik dalam pemasaran. *Customer Profiling* adalah menggunakan data untuk menggambarkan profil suatu segmen pelanggan. Sedangkan *Customer Segmentation* adalah tindakan yang membagi pelanggan menjadi suatu kelompok yang berbeda, dan biasanya tidak ada suatu pelanggan yang masuk dalam dua segmen. [8]

Ada beberapa cara yang dapat dilakukan dalam hal untuk melakukan *Customer Profiling and Segmentation*:

- Analisis RFM  
Salah satu tipe yang paling umum digunakan untuk melakukan *Customer Profiling and Segmentation* adalah RFM. Ini adalah metode yang digunakan untuk melakukan segmentasi berdasarkan kapan terakhir pelanggan melakukan transaksi, sering tidaknya melakukan transaksi dan transaksi yang dilakukan pada jumlah sedikit atau besar.
- Analisis Demografis  
Pada umumnya analisis demografis ini sangat erat kaitannya dengan geografis atau lokasi pelanggan itu berasal. Namun dalam beberapa penelitian demografis ini juga dapat diartikan melakukan segmentasi berdasarkan umur, jenis kelamin, pendapatan, dan status perkawinan.
- Analisis *Life Stage*  
Analisis *Life Stage* lebih dekat kepada analisis *behavior* atau analisis kebiasaan. Analisis ini berkaitan dengan apakah pelanggan memiliki kebiasaan yang sama apabila umurnya remaja, atau bagaimana.

Pada tugas akhir ini akan dilakukan tiga analisis lain selain analisis dengan model RFM. Yaitu analisis geografis, analisis tipe pelanggan, dan analisis *behavior*.

### **2.2.2.1. Analisis Geografi**

Geografi telah menjadi komponen praktis yang penting dari strategi pemasaran. Hal ini didorong untuk sebagian besar oleh tujuan ekspansi organisasi yang memaksa manajer untuk mempertimbangkan tata ruang pengiriman dan iklan sistem semakin kompleks saat peluncuran dan manajemen produk baru. [9]

Pada langkah dengan tren ini, para peneliti di bidang pemasaran dan ekonomi telah mengembangkan minat dalam aspek spasial pertumbuhan dan struktur pasar. Tradisi penelitian yang dihasilkan telah disebut "geografi ekonomi baru." Aliran ini penelitian - yang dimulai pada tahun 1970 di bidang organisasi industry. [10]

Sadar akan pentingnya geografi dalam strategi pemasaran maka PT. XYZ juga memperlakukan pelanggannya secara berbeda jika letak pelanggan jaraknya jauh atau dekat dari PT. XYZ. Oleh karena itu akan dibuat map yang akan memperlihatkan persebaran pelanggan PT. XYZ pada setiap segmen agar dapat mengetahui bagaimana cara memperlakukan pelanggan dengan geografi yang berbeda. PT. XYZ berharap jika pelanggan yang jaraknya jauh dari PT. XYZ akan melakukan transaksi dalam jumlah yang banyak meskipun jarang melakukan transaksi.

### **2.2.2.2. Analisis Tipe Pelanggan**

Hal yang paling penting dalam menjalankan bisnis adalah memahami pelanggan, meskipun barang yang ditawarkan perusahaan sangat bagus tapi apabila pelanggan berpikir bahwa produk itu tidak bagus atau tidak diperlukan maka produk perusahaan tidak akan laku. [11]

Ada sepuluh cara bagaimana cara perusahaan dapat mengetahui atau memahami pelanggannya. Salah satunya adalah siapakah pelanggan perusahaan. Jika perusahaan menjual produk kepada individu, yang harus dicari adalah

jenis kelamin pelanggan, umur, status pernikahan, dan gaji untuk memahami pelanggan. Jika perusahaan menjual kepada perusahaan lain atau bisnis lain, maka perusahaan harus mencari tau seberapa besar dan bisnis apa yang menjadi pelanggan perusahaan. [11]

Karena itu PT. XYZ membagi pelanggannya menjadi empat kategori yaitu apotik, rumah sakit, perusahaan/yayasan, dan klinik. Penanganan tiap kategori pelanggan sudah dibedakan secara umum, contohnya mengutamakan pelanggan dengan tipe apotik dan rumah sakit karena biasanya melakukan transaksi yang sering dan dalam skala yang besar.

### **2.2.2.3. Analisis Behavior**

Mempelajari perilaku konsumen jelas menarik bagi pemasar, tetapi merupakan studi yang menarik dalam dirinya sendiri, bahkan untuk non-marketer, karena kita semua konsumen. Pada akhirnya, konsumen memegang semua kekuasaan di dunia bisnis - seperti Sam Walton, pendiri Walmart, terkenal mengatakan, "Hanya ada satu bos - pelanggan. Dan ia dapat memecat setiap orang di perusahaan, dari ketua di bawah, hanya dengan menghabiskan uangnya di tempat lain 'Walton selalu menganggap dirinya sebagai agen bagi para pelanggannya, menemukan mereka nilai terbaik untuk uang. Filosofi sederhana ini pindah Wal Mart \* dari satu toko kecil di Arkansas untuk menjadi pengecer terbesar di dunia dalam seumur hidup Walton. [12]

Ada banyak cara untuk memahami kebiasaan pelanggan, salah satunya adalah kebiasaan pelanggan dalam melakukan transaksi atau waktu pelanggan dalam melakukan transaksi. Untuk membantu analisis ini maka pada tugas akhir ini akan ditampilkan kebiasaan transaksi pelanggan per minggu. Karena PT. XYZ memerlukan laporan ini untuk mengetahui pada minggu berapa

pelanggan banyak melakukan transaksi, pada minggu dan bulan apa pelanggan melakukan banyak atau sedikit transaksi.

### 2.2.3. Model RFM

Konsep dari RFM pertama diperkenalkan oleh Bult dan Wansbeek pada tahun 1995 [13] dan dibuktikan bahwa RFM adalah konsep yang paling efektif ketika diaplikasikan pada data transaksi oleh Blattberg. [14] Kepanjangan dari RFM adalah *Recency*, *Frequency*, dan *Monetary*. *Recency* berarti hari analisis dikurangi dengan hari terakhir saat pelanggan melakukan transaksi. *Frequency* berarti seberapa sering seorang pelanggan melakukan transaksi. Dan yang terakhir *monetary* adalah jumlah uang yang telah dihabiskan oleh seorang pelanggan.

**Tabel 2.2 Contoh Data RFM Pelanggan**

Customer Id	<i>Recency</i> (Hari)	<i>Frequency</i> (Jumlah)	<i>Monetary</i> (Ribu Rupiah)
1	3	6	540
2	6	10	940
3	45	1	30
4	21	2	64
5	14	4	169
6	32	2	55
7	5	3	130
8	50	1	950
9	33	15	2430



Customer Id	<i>Recency</i> (Hari)	<i>Frequency</i> (Jumlah)	<i>Monetary</i> (Ribu Rupiah)
10	10	5	190

Pada model RFM, *Recency* yang mengacu pada interval hari terakhir dimana pelanggan melakukan transaksi, dipercaya oleh banyak penjual bahwa pembeli yang paling baru melakukan transaksi akan cenderung untuk membeli lagi daripada pelanggan yang melakukan transaksi lebih lama. Untuk *Frequency* yang merepresentasikan jumlah transaksi pada jangka waktu tertentu, banyak penjual yang mengasumsikan bahwa pelanggan yang membuat transaksi lebih banyak akan lebih cenderung untuk melakukan transaksi pembelian lagi daripada pelanggan yang melakukan transaksi lebih sedikit. Sedangkan untuk *Monetary* yang merepresentasikan uang yang telah dihabiskan oleh pelanggan dalam jangka waktu tertentu, ini juga diasumsikan bahwa pelanggan yang menghabiskan lebih banyak uang daripada pelanggan lain akan melakukan pembelian atau penggunaan uang yang lebih. [15]

Pada penelitian yang lainnya banyak perubahan pada analisis RFM. Yang paling sering digunakan adalah *Weighted RFM* atau *WRFM* yang menggunakan model *AHP* untuk memberikan poin mana dari *Recency*, *Frequency*, dan *Monetary* yang lebih penting. Masih banyak lagi penelitian yang menggunakan perubahan dari model RFM.

Pada model RFM ini juga dapat digunakan setelah melakukan *clustering* yaitu dengan mencari rata-rata masing-masing RFM untuk setiap *cluster*. Kemudian dari rata-rata tersebut kita dapat menentukan bahwa *cluster* tersebut termasuk pada  $R \uparrow F \uparrow M \uparrow$  atau termasuk pada yang lain.

#### 2.2.4. Normalisasi min-max

Normalisasi adalah proses untuk menyamakan tipe data. Normalisasi data ini dilakukan karena perbedaan skala yang

terjadi pada masing-masing *Recency*, *Frequency*, dan *Monetary* terlalu besar. Misalnya pada *Monetary* angka terkecil adalah satu juta, sedangkan pada *Frequency* dan *Recency* angka paling kecil menunjukkan angka 1. Oleh karena itu dibutuhkan normalisasi agar jarak di antara ketiga variabel itu tidak terlalu jauh. Normalisasi RFM pada tugas akhir ini menggunakan metode Min-Max. Untuk nilai  $R_{ci}$ ,  $F_{ci}$ , dan  $M_{ci}$  pada persamaan 2.1, 2.2, dan 2.3 akan didefinisikan sebagai berikut:

$$R_{ci} = \frac{O_i^R - \min^R}{\max^R - \min^R} (\text{newmax}^R - \text{newmin}^R) + \text{newmin}^R \quad 2.1$$

$$F_{ci} = \frac{O_i^F - \min^F}{\max^F - \min^F} (\text{newmax}^F - \text{newmin}^F) + \text{newmin}^F \quad 2.2$$

$$M_{ci} = \frac{O_i^M - \min^M}{\max^M - \min^M} (\text{newmax}^M - \text{newmin}^M) + \text{newmin}^M \quad 2.3$$

$O_i^R$ ,  $O_i^F$ , dan  $O_i^M$  adalah nilai mentah yang akan dinormalisasi untuk masing-masing *Recency*, *Frequency*, dan *Monetary*. Sedangkan untuk  $\min^R$ ,  $\min^F$ , dan  $\min^M$  adalah nilai minimum dari masing-masing R, F, dan M. Begitu pula  $\max^R$ ,  $\max^F$ , dan  $\max^M$  yang menunjukkan nilai maksimal dari R, F, dan M.

$\text{Newmax}^R$ ,  $\text{newmax}^F$ , dan  $\text{newmax}^M$  adalah rentang nilai maksimal R, F, dan M baru yang mempunyai nilai 1, sedangkan  $\text{newmin}^R$ ,  $\text{newmin}^F$ , dan  $\text{newmin}^M$  adalah rentang baru nilai minimal masing-masing R, F, dan M yang bernilai 0. Hasil dari persamaan 2.1, 2.2, dan 2.3 akan mempunyai nilai yang rentangnya 0 sampai 1.

### 2.2.5. Penggalan Data

Penggalan data adalah terminologi yang digunakan untuk mendeskripsikan penemuan atau penggalan knowledge dari jumlah data yang besar. Secara teknis, penggalan data merupakan sebuah proses yang menggunakan teknik statistika, matematika, dan kecerdasan buatan untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang berpola dari himpunan data dengan ukuran besar. [16]

Secara skematis Goronescu [17] membagi langkah proses pelaksanaan penggalian data ke dalam tiga aktivitas yaitu:

- Eksplorasi data, terdiri dari aktivitas pembersihan data, transformasi data, pengurangan dimensi, pemilihan ciri, dan lain-lain.
- Membuat model dan pengujian validitas model, merupakan pemilihan terhadap model-model yang sudah dikembangkan yang cocok dengan kasus yang dihadapi. Dengan kata lain, dilakukan pemilihan model secara kompetitif
- Penerapan model dengan data baru untuk menghasilkan perkiraan dari kasus yang ada. Tahap ini merupakan tahap yang menentukan apakah model yang telah dibangun dapat menjawab permasalahan yang dihadapi.

Penggalian data telah menjadi alat yang populer dalam menyelesaikan permasalahan bisnis yang rumit di berbagai area. Tujuan penggalian data adalah untuk menyelesaikan permasalahan atau mengeksplorasi kesempatan bisnis yang muncul untuk menciptakan keuntungan kompetitif yang berkelanjutan. Berikut merupakan aplikasi penggalian data dalam bisnis:

- Pengelolaan hubungan pelanggan
- Perbankan
- Penjualan dan logistic
- Manufaktur dan produksi
- Asuransi
- Perangkat keras dan perangkat lunak computer
- Pemerintahan
- Industri agen perjalanan
- Kesehatan
- Olahraga [18]

### **2.2.6. Clustering**

*Clustering* atau dalam Bahasa Indonesia dikenal sebagai pengelompokkan adalah pengelompokkan beberapa objek yang

mempunyai kesamaan atau kemiripan karakter. Berbeda dengan klafikasi, pengelompokkan ini belum diketahui label kelas pada pengelompokannya. Teknik pengelompokan ini bertujuan untuk mengetahui objek mana saja yang ada pada satu segmen, dan apa perbedaan karakteristik tiap segmen. *Fuzzy C-Means* adalah metode pengelompokkan yang membolehkan sebuah data menjadi anggota di lebih dari satu kelompok, atau biasa disebut pengelompokkan dengan kategori tumpang tindih. [19]

#### **2.2.6.1. Algoritma Fuzzy C-Means**

Pengelompokan dengan metode Fuzzy C-Means (FCM) didasarkan pada teori logika fuzzy. Teori ini pertama kali dikenalkan oleh Lofi Zadeh (1965) dengan nama himpunan fuzzy (fuzzy set). Dalam teori fuzzy, keanggotaan sebuah data tidak diberikan nilai secara tegas dengan nilai 1 (menjadi anggota) atau 0 (bukan anggota), melainkan dengan derajat keanggotaan yang jangkauan nilainya dari 0 sampai 1. Nilai keanggotaan suatu data dalam sebuah himpunan menjadi 0 ketika sama sekali tidak menjadi anggota, dan menjadi 1 ketika menjadi anggota secara penuh dalam suatu himpunan. Umumnya nilai keanggotaannya antara 0 dan 1. Semakin tinggi nilai keanggotaannya, maka semakin tinggi derajat keanggotaannya. Begitu pula sebaliknya. Kaitannya dengan K-Means, sebenarnya FCM merupakan versi fuzzy dari K-Means dengan beberapa modifikasi yang membedakannya dengan K-Means. [20]

Pada tugas akhir ini digunakan algoritma FCM karena memiliki nilai entropy dan purity yang lebih baik dibanding dengan K-Means. Entropy dan purity adalah dua jenis pengukuran validasi eksternal. Pengukuran validasi eksternal ini dilakukan untuk mengukur kualitas klaster yang terbentuk dari dua algoritma di atas yaitu FCM dan K-Means. Purity adalah salah satu jenis pengukuran validasi yang paling utama untuk menentukan kualitas klaster. Purity dari klaster diukur berdasarkan label kelas atau ground thruth yang disebut

entropy. Semakin rendah entropy berarti lebih baik. Entropy juga digunakan untuk menemukan jumlah gangguan. Purity diukur dengan rumus 2.4 dan entropy diukur dengan rumus 2.5, berikut rumusnya:

$$Purity = \sum_i p_i (\max_j (p_{ij}/p_i)) \quad 2.4$$

Dimana nilai-nilai dari  $p_{ij}$ ,  $p_i$ , dan  $p_j$  dihitung berhubungan dengan label kelas.

$$Entropy = \sum_i p_i (\sum_j (p_{ij}/p_i) \log(p_{ij}/p_i)) \quad 2.5$$

Dimana nilai-nilai dari  $p_{ij}$ ,  $p_i$ , dan  $p_j$  dihitung berhubungan dengan label kelas.

### 2.2.6.2. Langkah-langkah Fuzzy C-Means

Berdasarkan konsep di atas, berikut proses pengelompokan dengan FCM yang dikembangkan oleh Bezdek [21]:

- Menentukan jumlah c klaster.
- Mengalokasikan data sesuai dengan jumlah klaster yang ditentukan
- Menghitung nilai titik tengah ( $V_{ij}$ ) dari masing-masing klaster, dengan persamaan 2.6 sebagai berikut:

$$V_{ij} = \frac{\sum_{k=1}^N (U_{ik})^M x_{kj}}{\sum_{k=1}^N (U_{ik})^M} \quad 2.6$$

Dimana:

$N$  = Jumlah data

$m$  = Pembobotan eksponensial

$U_{ik}$  = Membership function data ke- $k$  ke klaster ke- $i$

- Menghitung nilai fungsi keanggotaan masing-masing data ke masing-masing c klaster, dengan persamaan 2.7 sebagai berikut:

$$U_{ik} = \sum_{j=1}^c \left( \frac{D(x_k, v_i)}{D(x_k, v_j)} \right)^{\frac{2}{m-1}} \quad 2.7$$

Dimana:

$U_{ik}$  = Membership function data ke-k ke klaster ke-i

$v_i$  = Nilai centroid klaster ke-i

$m$  = Pembobotan eksponensial

$x$  = Titik data

Tingkat kemungkinan yang lebih tinggi ke suatu kelompok akan mempunyai nilai fungsi keanggotaan ke kelompok tersebut yang mendekati angka 1 dan ke kelompok yang lain mendekati angka 0.

- e. Kembali ke langkah c. Jika perubahan nilai fungsi keanggotaan lebih dari nilai ambang batas yang ditentukan atau jika perubahan pada nilai titik tengah lebih dari nilai ambang batas yang ditentukan, atau jika perubahan pada nilai fungsi obyektif lebih dari nilai ambang batas yang ditentukan maka nilai fungsi obyektif didefinisikan dalam persamaan 2.8:

$$J(U, V) = \sum_{k=1}^N \sum_{i=1}^c u_{ik}^m D(x_k, v_i)^m \quad 2.8$$

Dimana:

$N$  = Jumlah data

$c$  = Jumlah klaster

$u_{ik}$  = Fungsi keanggotaan data ke-k ke klaster ke-i dengan nilai antara 0 – 1

$v_i$  = Nilai titik tengah ke-i

### 2.2.7. Validasi

Validasi adalah proses menilai seberapa baik model *mining* terhadap data asli yang digunakan. Hal ini dilakukan untuk memahami kualitas dan karakteristik dari model mining yang kita gunakan sebelum menetapkan dalam lingkungan produksi. [22]. Sedangkan validasi *cluster* berkaitan dengan kualitas *cluster* yang dihasilkan oleh algoritma untuk *clustering* data.

Ketika informasi mengenai keanggotaan kelas yang benar tersedia, maka validasi *cluster* eksternal dapat digunakan, validasi eksternal menyediakan cara yang objektif untuk menilai kinerja algoritma. Namun ketika pengetahuan eksternal tidak tersedia maka langkah-langkah validasi internal yang digunakan, dimana dengan menggunakan validasi internal akan mengukur kualitas *cluster* berdasarkan sifat instrinsik dari data. [23]

### 2.2.7.1. Sum of Square Error (SSE)

SSE adalah suatu cara atau metode yang paling sederhana dan paling banyak digunakan untuk mengukur kriteria dalam evaluasi pengelompokan [24]. SSE ini hanya menggunakan informasi dari data itu sendiri dan tidak menggunakan informasi apapun dari luar. SSE juga mengukur kekompakan kelompok menggunakan beberapa ukuran. Maka dari itu SSE masuk ke dalam *internal quality criteria*. Rumus untuk mencari SSE terdapat pada persamaan 2.9:

$$SSE = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad 2.9$$

Dimana:

$C_k$  = himpunan contoh dalam pengelompokan k

$\mu_k$  = rata-rata vektor dari pengelompokan k

Berikut cara untuk menghitung rata-rata vektor dari pengelompokan k ( $\mu_k$ ) dengan persamaan 2.10:

$$\mu_{kj} = \frac{1}{N_k} \sum_{x_i \in C_k} x_{ij} \quad 2.10$$

Dimana:

$N_k = |C_k|$  = jumlah contoh milik pengelompokan k. [24]

### 2.2.7.2. Davies Bouldin Index (DBI) atau Index Dunn

Davies Bouldin Index adalah metode validasi yang ide dasarnya untuk partisi yang bagus, pemisahan *inter cluster* maupun homogenitas *intracluster* dan *compactness* harus tinggi. Sedangkan Indeks Dunn diperkenalkan oleh J.C Dunn (1973) sebagai metrik untuk validitas hasil kelompok.

Validitas kelompok ditentukan dengan menggunakan diameter kelompok (kohesi) dan jarak antara dua kelompok (separasi) Indeks Dunn merupakan indeks yang umum digunakan untuk memvalidasi keoptimalan kelompok yang dihasilkan dari algoritma lain, seperti algoritma K-Means. Karena Index Dunn lebih dekat kepada *clustering kmeans*, maka paa tugas akhir ini akan membandingkan hasil SSE dan DBI.

Untuk menentukan ideks DB, perlu mendefinisikan ukuran dipersi dan ukuran kesamaan *cluster*. Selain itu DBI digunakan untuk mengukur jarak antara *cluster*  $C_i$  dan  $C_j$ , pada saat yang sama digunakan untuk meminimalkan jarak antara titik-titik dalam *cluster* degan pusat *cluster*, jarak  $s_c(Q_k)$  dalam *cluster*  $Q_k$  didefinisikan pada rumus 2.11:

$$s_c(Q_k) = \sum_i \frac{\|X_i - C_k\|}{N_k} \quad 2.11$$

Dimana  $N_k$  adalah nomer dari titik yang dimiliki oleh *cluster*  $Q_k$  dan  $C_k = \frac{1}{N_k} \sum x_i$  Jarak antara *cluster* didefinisikan dengan  $d_{cc} = \|C_k - C_l\|$  sehingga DB index bisa didefinisikan sebagai berikut:

$$DB(nc) = \frac{1}{nc} \sum_{k=1}^{nc} \max l \neq m \left\{ \frac{s_c(Q_k) + s_c(Q_l)}{d_{cc}(Q_k, Q_l)} \right\} [25] [26]$$

### 2.2.8. Aplikasi R untuk Proses Clustering, Validasi, dan Visualisasi

R adalah software aplikasi yang memfasilitasi pengguna untuk memanipulasi data, mengkalkulasi data dan menyajikan data dengan grafis. Keuntungan yang didapat dalam penggunaan aplikasi R adalah:

- Efektif untuk menangani data dan memiliki penyimpanan sendiri
- Memiliki perhitungan array dan matriks tertentu
- Fasilitas untuk memvisualisasikan informasi untuk analisis data



- Tools yang baik dalam develop karena sederhana dan punya bahasa pemrograman yang efektif. [27]

Satu lagi keuntungan yang didapat adalah karena aplikasi R ini *open source*. Karena keuntungan yang didapat pada aplikasi R tersebut, maka aplikasi R akan digunakan pada tugas akhir ini saat membentuk klaster dengan algoritma Fuzzy C-Means dan visualisasi dengan menggunakan jenis visualisasi *scatter plot*. Pada aplikasi R, terdapat *package* khusus untuk algoritma FCM dan akan digunakan *package* ggplot yang terdapat di aplikasi R untuk visualisasinya.

Untuk jenis visualisasinya, akan digunakan jenis visualisasi *scatter plot*. Jenis visualisasi *scatter plot* memberikan visualisasi data untuk melihat tren, kelompok, dan persebaran data sehingga mempermudah untuk identifikasi hubungan antar variable [28]. *Scatter plot* juga dapat digunakan untuk memberikan visualisasi terhadap data yang mempunyai variabel pengukur sebanyak dua buah atau lebih [29]. Pada tugas akhir ini, jelas variabel pengukurnya terdapat tiga buah yaitu model RFM; *Recency*, *Frequency*, dan *Monetary*. Karena hal itu jenis visualisasi *scatter plot* digunakan pada tugas akhir kali ini. Setelah dibuatnya visualisasi data menggunakan jenis visualisasi *scatter plot*, diharapkan dapat membantu analisis mengenai data dan juga memperluas analisis tentang data transaksi ini.

### 2.2.8.1. Proses Clustering

Proses *clustering* pada aplikasi R menggunakan *package* e1071 dengan fungsi *cmeans* di dalamnya. Karena hanya pada *package* ini terdapat fungsi *clustering* dengan metode *Fuzzy C-Means Clustering*. Contoh *coding* untuk *cmeans* adalah sebagai berikut:

```
cmeans(x, centers, iter.max = 100, verbose =  
FALSE, dist = "21ank mean", method =  
"cmeans", m = 2, rate.par = NULL, weights = 1,  
control = list())
```

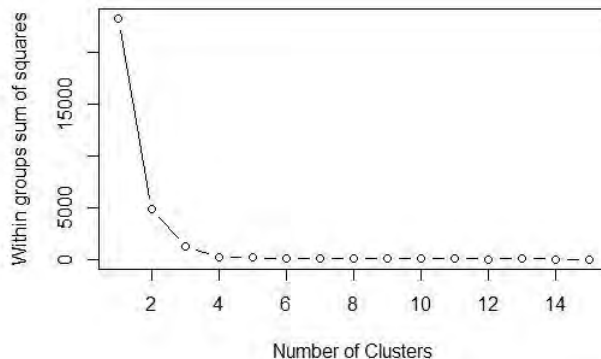
Dimana  $x$  adalah data yang akan diolah, *centers* adalah jumlah *center* yang akan dibentuk pada proses *clustering* dan  $m$  adalah jumlah variabel yang akan ditinjau untuk membentuk kluster. [30]

### 2.2.8.2. Proses Validasi

Proses validasi ini menggunakan rumus normal dan memasukkannya pada aplikasi R. Kemudian akan digambarkan menjadi plot agar mudah memahaminya. Contoh *coding* untuk SSE adalah sebagai berikut.

```
wss <- (nrow(mydata)-1)*sum(apply(mydata,2,var))
  for (I in 2:15) wss[i] <- sum(kmeans(mydata,
centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of
Clusters",
     ylab="Within groups sum of squares")
```

Dimana *mydata* adalah data yang ingin divalidasi 22ank means adalah metode *clustering* yang dilakukan. Hasil dari proses ini adalah gambar grafik yang contohnya dapat dilihat pada gambar 2.1.

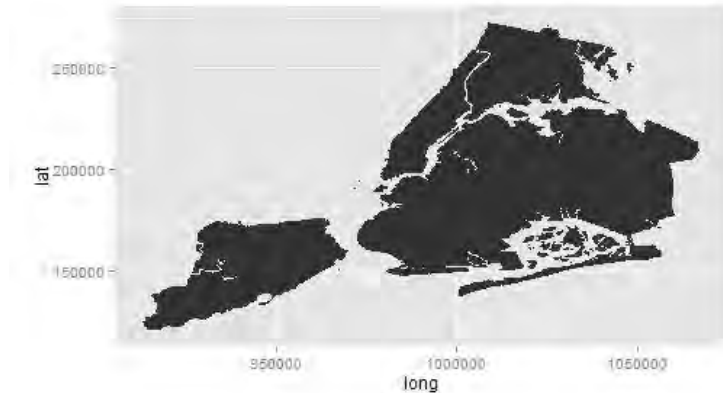


**Gambar 2.1 Contoh Hasil Validasi SSE**

Dengan melihat grafik di atas maka dapat diambil kesimpulan bahwa *cluster* maksimal adalah 4. Karena diambil dari SSE yang mulai stabil. [31]

### 2.2.8.3. Proses Visualisasi

Proses visualisasi ini akan menggunakan tiga tipe visualisasi yaitu visualisasi map dan *heat map*. Untuk visualisasi map akan digunakan *package* googleVis untuk menunjukkan lokasi dari masing-masing pelanggan. Sebenarnya selain *package* googleVis ini ada *package* lain yang dapat menampilkan lokasi pelanggan juga. Seperti RgoogleMaps dan rgdal dengan ggplot2. Namun pada rgdal dan ggplot2 input mapnya harus berbentuk shp. SHP adalah tipe map dengan membuat partisi-partisi map seperti gambar 2.2.



**Gambar 2.2 Contoh Map dengan SHP**

Hal ini tidak dibutuhkan pada tugas akhir ini karena tugas akhir ini hanya ingin menunjukkan lokasi pelanggan tanpa membagi wilayahnya. Sedangkan RgoogleMaps lebih susah digunakan karena harus menyertakan skala mapnya. Untuk

*package* googleVis lebih mudah dibuat dan dapat digantikan skala mapnya. Untuk contoh *coding* pada aplikasi R adalah sebagai berikut.

```
gvisMap(data, locationvar = "",
tipvar = "", options = list(),
```

~~Diharapkan~~ *data*) adalah data yang akan dilakukan analisis dan *locationvar* adalah *longitude* dan *latitude* dari masing-masing pelanggan, dan *tipvar* adalah atribut tambahan yang ingin ditampilkan. [32]

Untuk visualisasi *heat map* akan digunakan *package* ggplot yang ada pada aplikasi R. *Package* ini digunakan karena hanya pada *package* ini visualisasi dapat dibuat dengan mudah dan bagus.

Visualisasi terakhir yang digunakan adalah *scatter plot*. *Scatter plot* merupakan bentuk visualisasi data yang efektif untuk melihat tren, konsentrasi dan persebaran data sehingga mudah dalam mengidentifikasi hubungan tiap variabel [28]. *Scatter plot* dapat digunakan untuk data-data yang mempunyai nilai numerik serta bersifat *multivariate* yaitu kumpulan data yang terdiri dari 2 variabel pengukur atau lebih [29]. Pembuatan *scatter plot* ini akan memanfaatkan *package* scatterplot3D. Pada tugas akhir ini akan menggunakan 3 variabel yaitu *recency*, *frequency*, dan *monetary*.

### **BAB III**

#### **METODOLOGI Pengerjaan Tugas Akhir**

Pada bagian ini akan dijelaskan mengenai proses pengerjaan tugas akhir. Penjelasan akan diberikan dalam bentuk flowchart yang meliputi studi literatur, pengumpulan data dan *preprocessing* data, proses *clustering* dengan algoritma Fuzzy C-Means, validasi hasil *clustering* dengan SSE, pembuatan visualisasi dan segmentasi hasil *clustering* pelanggan, dan penyusunan buku tugas akhir. Flowchart dari alur pengerjaan akan ditunjukkan oleh gambar 3.1.

#### **3.1. Studi Literatur**

Pada tahap studi literatur dilakukan pemahaman terhadap masalah yang terjadi pada studi kasus dan semua referensi mengenai penyelesaian masalah yang ada pada studi kasus. Studi literatur ini didapatkan dari *e-book*, *paper*, dan informasi yang ada pada *website* pemerintah. Referensi tersebut terkait *trend* bisnis produsen obat, penggalian data, analisis RFM, segmentasi pelanggan, algoritma *Fuzzy C-Means*, SSE, visualisasi, dan aplikasi R.

#### **3.2. Pengumpulan Data dan *Preprocessing* Data**

Langkah selanjutnya adalah pengumpulan data. Data yang dikumpulkan adalah data rekap transaksi per hari yang ada pada PT. XYZ. Data transaksi dimulai dari bulan Januari 2013 sampai bulan November 2015, data tersebut berupa tiga buah *file excel*. Masing-masing *file excel* tersebut memiliki atribut tanggal, bulan, tahun, KD PDU, kode dokter, kode langganan, rayon, kode area, nama area, nama langganan, alamat langganan, RSM, T/B/L, AP 1/2, A/B, DM, line, mr, kode produk, nama produk, unit, hna pusat, tot by pusat, hna msd, tot by msd, produk grup, kode salur, kode bonus, dan produk kontes. Setelah mendapatkan data, maka proses selanjutnya adalah merubah data tersebut menjadi data yang siap olah yang disebut *preprocessing* data. Untuk lebih rincinya,

*preprocessing* data dilakukan dalam beberapa tahap yaitu, pembersihan data, transformasi data, dan normalisasi data.

a. Pembersihan data

Proses pembersihan data ini dilakukan agar atribut yang ada pada data yang didapat pada pengumpulan data tidak mengganggu pada saat pengerjaan karena jumlah atribut yang sangat banyak. Oleh karena itu pada pembersihan data ini dilakukan penghapusan atribut yang tidak diperlukan saat mengerjakan tugas akhir ini.

b. Transformasi data

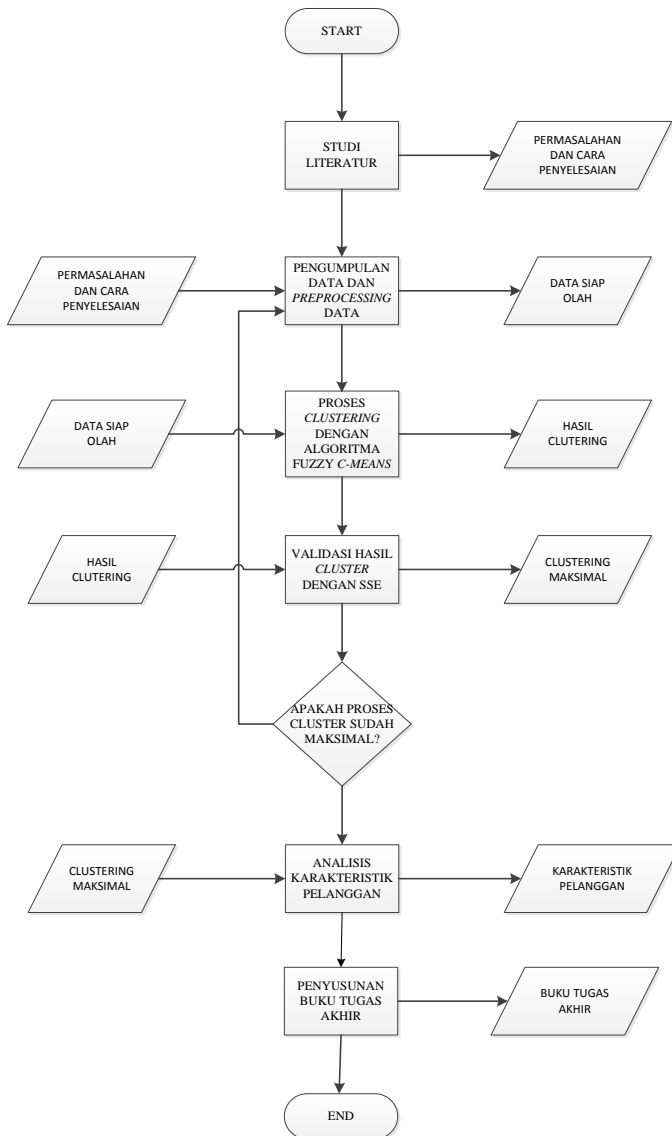
Setelah data tersebut dibersihkan, langkah selanjutnya adalah membuat data tersebut untuk menjadi model RFM. Untuk mempermudah aktivitas ini maka data yang sudah dibersihkan akan dimasukkan ke dalam SQL Server Management Studio. Lalu akan dibentuk model RFM pada SQL Server Management Studio yang didukung dengan Microsoft Excel.

c. Normalisasi data

Setelah mendapatkan RFM dari setiap pelanggan, langkah selanjutnya adalah normalisasi RFM. Normalisasi RFM ini dilakukan sesuai pada bab 2.2.3 dan lebih tepatnya normalisasi dilakukan dengan menggunakan persamaan 2.1, 2.2, dan 2.3 untuk masing-masing *Recency*, *Frequency*, dan *Monetary*.

### **3.3. Proses *Clustering* dengan Algoritma Fuzzy C-Means**

Pada tahap ini dilakukan proses klasterisasi atau pengelompokkan pelanggan dengan menggunakan metode *Fuzzy C-Means*. Data yang digunakan dalam tahapan ini adalah data yang sudah dinormalisasi. Setelah itu dilakukan klasterisasi yang tahapnya sudah dijelaskan pada bab 2.5.1 dan 2.5.2.



**Gambar 3.1 Metodologi Penelitian**

### **3.4. Validasi Hasil *Cluster* dengan SSE**

Setelah mengetahui hasil klasterisasi yang telah digunakan, maka langkah selanjutnya adalah uji validasi dengan SSE. SSE digunakan untuk menganalisis variasi pada jarak antar pusat klaster yang terbentuk. Tahapan ini lebih jelasnya dijelaskan pada 2.6.

### **3.5. Analisis Karakteristik Pelanggan**

Setelah mendapatkan hasil klaster yang terbaik maka langkah selanjutnya yang dilakukan adalah melakukan analisis terhadap klaster yang terbentuk. Analisis yang akan dilakukan adalah analisis geografis, analisis demografi, analisis *behavior* dan analisis masing-masing atribut RFM. Proses ini sudah dijelaskan pada dasar teori 2.2. Kemudian akan dihasilkan karakteristik pelanggan dari perusahaan.

### **3.6. Penyusunan Buku Tugas Akhir**

Tahap terakhir dari penelitian ini adalah menyusun buku tugas akhir yang akan digunakan sebagai bahan evaluasi penelitian. Penyusunan buku tugas akhir ini disesuaikan dengan standar / template yang berlaku di Institusi.



## **BAB IV**

### **PERANCANGAN**

Pada bab ini akan dijelaskan bagaimana rancangan dari penelitian tugas akhir yang meliputi subyek dan obyek dari penelitian, pemilihan subyek dan obyek penelitian dan bagaimana penelitian akan dilakukan.

#### **4.1. Pengumpulan Data**

Pengumpulan data adalah proses yang utama pada tugas akhir kali ini. Proses ini menentukan keberhasilan dari tugas akhir ini karena data yang digunakan pada saat segmentasi dan analisis akan didapatkan pada proses kali ini. Pengumpulan data ini dilakukan untuk mendapatkan data mentah yang selanjutnya akan diproses dalam segmentasi dan analisis. Pada pengumpulan data yang diminta adalah data transaksi harian yang ada pada PT. XYZ.

Data mentah yang didapatkan berupa tiga buah Ms.Excel yang di dalamnya terdapat rekap transaksi harian dari Januari 2013 sampai ovember 2014. Data mentah ini memiliki atribut sebagai berikut yang terlihat pada tabel 4.1

**Tabel 4.1 Atribut dan Tipe Data Mentah**

<b>Atribut</b>	<b>Tipe data pada Excel</b>	<b>Keterangan</b>
Bulan	Number	Bulan transaksi
Tahun	Number	Tahun transaks
KODE LANGG	Text	Kode langganan dari pelanggan
TGL	Number	Tanggal transaksi
RAYON	Text	Daerah pelanggan

<b>Atribut</b>	<b>Tipe data pada Excel</b>	<b>Keterangan</b>
KODE AREA	Text	Kode area dari pelanggan
NAMA AREA	Text	Nama area dari pelanggan
NAMA LANGGANAN	Text	Nama dari pelanggan
ALMTLANG	Text	Alamat dari pelanggan
KODE PROD	Text	Kode produk yang dibeli pelanggan
NAMA PROD	Text	Nama produk yang dibeli pelanggan
UNIT	Number	Jumlah produk yang dibeli pelanggan
hna pusat	Number	Harga satuan produk
Tot by pusat	Number	Total harga yang harus dibayar
PRODUK GROUP	Text	Grup produk dari yang dibeli pelanggan

Berdasarkan atribut di atas maka didapatkan keseluruhan data transaksi yang berjumlah 141693 transaksi dalam 23 bulan terakhir pada tahun 2013 sampai 2014. Terdapat beberapa transaksi yang disimpan pada Ms.Excel seperti yang terlihat pada tabel 4.2.

**Tabel 4.2 Cuplikan Data Mentah**

<b>BLN</b>	<b>THN</b>	<b>KODE LANGG</b>	<b>TGL</b>	<b>KODE AREA</b>	<b>KODE PROD</b>	<b>UNIT</b>	<b>hna pusat</b>	<b>tot by pusat</b>	<b>PRODUK GROUP</b>
<b>01</b>	2013	95854	3	101	TKNXK	1	234000	234000	KALNEX
<b>01</b>	2013	95854	3	101	TZEGA	3	100000	300000	ZEGAVIT
<b>01</b>	2013	95854	4	101	AKMEB	10	26500	265000	KALMETHASONE
<b>01</b>	2013	95854	4	101	AKMED	20	25000	500000	KALMETHASONE
<b>01</b>	2013	95799	5	102	VCRVD	6	290000	1740000	CRAVIT
<b>01</b>	2013	95854	7	101	KCLMB	1	309000	309000	CLIMADAN
<b>01</b>	2013	95854	7	101	TKNXK	2	234000	468000	KALNEX
<b>01</b>	2013	95854	7	101	TKTFA	2	72000	144000	KALTROFEN
<b>01</b>	2013	95854	7	101	TKTFB	2	125000	250000	KALTROFEN

## 4.2. Praprocessing Data

Setelah mendapatkan data mentah yang telah dilakukan pada proses sebelumnya, maka langkah selanjutnya adalah penyiapan data tersebut agar siap untuk diolah dan dilakukan segmentasi. Langkah pertama adalah menghilangkan atribut-atribut yang tidak dibutuhkan dalam penghitungan RFM. Proses ini menghasilkan tabel yang lebih kecil dengan atribut yang hanya dibutuhkan untuk menghitung RFM. Contoh data dari hasil penyeleksian atribut bisa dilihat pada tabel 4.3

**Tabel 4.3 Hasil Penyeleksian Atribut**

TGL TRANS	KODE LANGGANAN	tot by pusat
3/1/2013	95854	234000
3/1/2013	95854	300000
4/1/2013	95854	265000
4/1/2013	95854	500000
5/1/2013	95799	1740000
7/1/2013	95854	309000
7/1/2013	95854	468000
7/1/2013	95854	144000
7/1/2013	95854	250000
7/1/2013	95854	200000
7/1/2013	95854	2500000
9/1/2013	95854	372000
9/1/2013	86787	234000

Setelah melakukan pembersihan atribut yang tidak berguna untuk dilakukannya *clustering*, maka selanjutnya adalah proses pembersihan data atau perbaikan data. Perbaikan data dilakukan terlebih dahulu, karena pada data mentah yang didapatkan dari PT. XYZ masih banyak redundansi data. Contohnya adalah

penggunaan nama langganan yang sama tetapi kode langganan berbeda atau menggunakan nama yang hanya berbeda spasi dan titik saja. Maka data tersebut akan diperbaiki dan menyamakannya dalam satu kode langganan dan nama langganan. Berikut cuplikan data yang perlu perbaikan ditunjukkan oleh tabel 4.4.

Selanjutnya adalah pembersihan data. Setelah melakukan perbaikan data yang ada, masih terdapat data yang memiliki nilai null di kode langganannya. Karena tidak mungkin lagi disamakan dengan kode langganan dan nama langganan lain karena tidak ada yang hampir sama, maka langkah terbaik yang diambil adalah penghapusan data. Berikut contoh data yang memiliki nilai null ditunjukkan pada tabel 4.4.

**Tabel 4.4 Contoh Pembersihan Data**

TGL TRANS	KODE LANGGANAN	tot by pusat
3/1/2013	95854	234000
3/1/2013	95854	300000
4/1/2013	95854	265000
4/1/2013	95854	500000
5/1/2013	95799	1740000
7/1/2013	95854	309000
7/1/2013	95854	468000
7/1/2013	95854	144000
7/1/2013	95854	250000
7/1/2013	95854	200000
7/1/2013	95854	2500000
9/1/2013	95854	372000
1/6/2013		375000

### 4.3. Mencari Nilai Atribut RFM

Setelah mendapatkan data yang sudah bersih dan siap untuk diolah lebih lanjut, maka langkah selanjutnya adalah mencari nilai *Recency*, *Frequency* dan *Monetary* untuk masing-masing dari pelanggan. Proses ini dilakukan pada SQL Server Management Studio, dengan menggunakan *query* yang sesuai untuk masing-masing atribut *Recency*, *Frequency*, dan *Monetary*.

Setelah didapatkan data yang lengkap dengan atribut yang sudah sesuai maka akan dilakukan tahap pencarian nilai R, F, dan M dengan menggunakan bantuan dari RDBMS (SQL Server). Dengan menggunakan skema database satu tabel yang berisi data transaksi yang telah diseleksi atributnya dan setelah pembersihan data seperti pada tabel 4.3 dan 4.4. Dengan menggunakan *query* yang sesuai untuk masing-masing atribut maka akan dihasilkan nilai RFM secara cepat. Penjelasan mengenai bagaimana mendapatkan nilai masing-masing atribut akan dijelaskan pada sub bab 4.3.1. hingga 4.3.4.

#### 4.3.1. Mencari Nilai *Recency* (R)

*Recency* adalah selisih hari antara tanggal dilakukannya analisis dan tanggal terakhir seorang pelanggan melakukan transaksi. Langkah pertama dalam mencari nilai *recency* adalah menentukan tanggal transaksi yaitu tanggal 11 Februari 2105. Jadi rumus umum sebenarnya adalah:

$$R = \text{Tanggal dilakukan analisis}(11-02-2015) - \text{tanggal terakhir pelanggan melakukan transaksi (dari Januari 2013 - November 2014)}$$

4.1

Pada SQL Server Management Studio yang pertama dilakukan adalah mencari tanggal terakhir dimana setiap customer melakukan transaksi. *Query* untuk mencari selisish tersebut terdapat pada sql query di bawah ini.

```
SELECT [KODE LANGGANAN], DATEDIFF(day, (max ([TGL
TRANS])) , '2015-02-11') AS Recency
```

```
from TA.dbo.PenjualanFix
group by [KODE LANGGANAN]
order by [KODE LANGGANAN]
```

#### 4.3.2. Mencari Nilai *Frequency* (F)

*Frequency* adalah jumlah transaksi yang dilakukan oleh pelanggan pada rentang waktu tertentu, pada kasus ini rentang waktunya yaitu Januari 2013 sampai November 2014. Pada SQL Management Studio dapat digunakan *query* COUNT dan DISTINCT untuk melakukan penghitungan transaksi yang dilakukan oleh tiap pelanggan. *Query*-nya dapat dilihat di bawah ini.

```
SELECT [KODE LANGGANAN], COUNT(DISTINCT([TGL
TRANS])) AS Freq
from TA.dbo.PenjualanFix
group by [KODE LANGGANAN]
order by [KODE LANGGANAN]
```

#### 4.3.3. Mencari Nilai *Monetary* (M)

*Monetary* adalah jumlah uang atau rupiah yang telah dihabiskan pelanggan pada rentang waktu tertentu, dalam kasus ini rentang waktunya yaitu Januari 2013 sampai November 2014. Pada SQL Management Studio, *monetary* dapat dicari dengan *query* SUM yang menjumlahkan tot\_pusat pada masing-masing pelanggan. *Query* pada SQL dapat dilihat.

```
SELECT DISTINCT [KODE LANGGANAN], SUM([tot by
pusat]) AS 'Monetary'
from TA.dbo.PenjualanFix
group by [KODE LANGGANAN]
order by [KODE LANGGANAN]
```

Ternyata setelah dilakukan pencarian nilai RFM dalam rentang waktu Januari 2013 sampai November 2014 maka diketahui ternyata pelanggan yang melakukan transaksi pada rentang waktu tersebut ada 475 pelanggan. Tabel 4.5 menunjukkan hasil *Recency*, *Frequency*, dan *Monetary*.

**Tabel 4.5 Cuplikan Nilai RFM Pelanggan**

Kode Langganan	Recency	Freq	Mon
9023594	749	2	200000
557173	537	2	566000
629472	586	2	500000
9023591	674	1	270000
531569	766	1	177000
614869	583	1	235000
571945	640	1	414000
620833	695	2	260000
86622	677	11	3245500
9023588	712	2	571000
549913	742	3	934000
86768	733	1	124000
9023587	543	2	200000

#### **4.4. Normalisasi RFM Menggunakan Metode min-max**

Setelah mendapatkan nilai dari masing-masing *Recency*, *Frequency*, dan *Monetary* selanjutnya yang perlu dilakukan adalah normalisasi dari masing-masing nilai RFM tersebut. Normalisasi ini dilakukan agar rentang perbedaan nilai di antara ketiga variabel tersebut tidak terlalu jauh. Karena normalisasi ini akan membuat nilai dari masing-masing RFM akan menjadi 0-1 saja. Nilai terendah dari masing-masing RFM akan berubah menjadi 0, begitu pula sebaliknya nilai tertinggi masing-masing RFM akan berubah menjadi 1. Rumus yang digunakan untuk melakukan normalisasi ada pada bab 2.2.



#### 4.4.1. *Recency*

Untuk perhitungan *Recency*, dengan mengadopsi rumus dari normalisasi tersebut kemudian diketahui nilainya seperti gambar di bawah ini.

**Tabel 4.6** Cuplikan Nilai *Recency* Pelanggan yang Telah Dinormalisasi

Kode Langganan	RecNorm
9023594	0.969609
557173	0.662808
629472	0.733719
9023591	0.861071
531569	0.994211
614869	0.729378
571945	0.811867
620833	0.891462
86622	0.865412
9023588	0.916064

#### 4.4.2. *Frequency*

Untuk perhitungan *Frequency*, dengan mengadopsi rumus dari normalisasi tersebut kemudian diketahui nilainya seperti gambar di bawah ini.

**Tabel 4.7** Cuplikan Nilai *Frequency* Pelanggan yang Telah Dinormalisasi

Kode Langganan	FreqNorm
9023594	0.001767
557173	0.001767
629472	0.001767
9023591	0
531569	0

Kode Langganan	FreqNorm
614869	0
571945	0
620833	0.001767
86622	0.017668

#### 4.4.3. *Monetary*

Untuk perhitungan *Monetary*, dengan mengadopsi rumus dari normalisasi tersebut kemudian diketahui nilainya seperti gambar di bawah ini.

**Tabel 4.8** Cuplikan Nilai *Monetary* Pelanggan yang Telah Dinormalisasi

Kode Langganan	MonNorm
9023594	8.63E-06
557173	2.7E-05
629472	2.37E-05
9023591	1.21E-05
531569	7.48E-06
614869	1.04E-05

## BAB V IMPLEMENTASI

Pada bab ini berisi tentang proses pengolahan data yang didapatkan dari tahap rancangan sebelumnya dan proses implementasi *data mining* dengan menggunakan metode *clustering*.

### 5.1. *Clustering* menggunakan Algoritma Fuzzy C-Means

Setelah mendapatkan nilai dari masing-masing RFM yang telah dinormalisasi, maka akan dilakukan proses *clustering* dengan algoritma Fuzzy C-Means. Karena belum diketahui berapa *cluster* yang maksimal maka pada tugas akhir ini dicoba dari jumlah *cluster* 3-8. Untuk proses *clustering* ini dilakukan dengan aplikasi R. Pada aplikasi R akan didapatkan setiap pelanggan dengan nilai kepemilikan pada setiap *cluster*, *centroid* dari setiap *cluster*, dan juga *cluster* terdekat dengan masing-masing pelanggan.

Pada aplikasi R akan digunakan fungsi `cmeans` yang ada pada *package* `e1071` yang sudah dijelaskan pada bab 2.2.7. 1. Berikut langkah-langkah untuk melakukan proses *clustering* pada aplikasi R.

- Langkah pertama untuk melakukan *clustering* dengan *tools* aplikasi R adalah mengunggah data ke aplikasi R, langkah ini akan ditunjukkan oleh *command* berikut. Cuplikan data RFMNorm dapat dilihat pada tabel 5.1.  
`dataku <- read.csv("RFMNorm.csv", header=T)`

**Tabel 5.1 Cuplikan Nilai RFM yang Telah Dinormalisasi**

Kode Langganan	Rec	Freq	Mon
9023594	0.004342	0.14841	0.000916
557173	0.011577	0.243816	0.001316
629472	0.001447	0.226148	0.002157

Kode Langganan	Rec	Freq	Mon
9023591	0.023155	0.111307	0.000652
531569	0.004342	0.084806	0.000361
614869	0.017366	0.160777	0.000851
571945	0.017366	0.100707	0.000509
620833	0.013025	0.136042	0.00248
86622	0.002894	0.120141	0.000636
9023588	0.001447	0.088339	0.000916
549913	0.002894	0.194346	0.002888
86768	0.007236	0.120141	0.001125
9023587	0.001447	0.14841	0.000721

- Kemudian melakukan pengaktifan *package* *e1071* yang di dalamnya terdapat fungsi *cmeans*.  

```
library("e1071", lib.loc="C:/Program Files/R/R-3.1.2/library")
```
- Langkah selanjutnya adalah proses inti dari *clustering* dengan fungsi *cmeans*. Dataku adalah data set yang sudah dibuat pada langkah pertama pada proses ini. Kemudian 6 adalah jumlah klaster yang ingin dibentuk. Dengan iterasi maksimal yaitu 100. Pada proses ini karena jumlah klaster maksimal belum diketahui maka angka 6 diganti mulai 3 sampai 8.  

```
cmeans(dataku, 6, iter.max = 100, verbose = FALSE, dist = "euclidean", method = "cmeans", m = 3, rate.par = NULL, weights = 1, control = list())
```

Hasil dari proses ini dapat dilihat pada tabel 5.1. Pada tabel 5.1 diperlihatkan untuk jumlah *cluster* 3 dan *centroid* dari tiap RFM pada masing-masing *cluster*. Untuk tabel 5.2, 5.3, 5.4, 5.5, dan 5.6 menunjukkan jumlah *cluster* dari 4-8.

Tabel 5.2 *Centroid* untuk C=3

Center	Rec	Freq	Mon
<i>Cluster</i>			
1	0.01345	0.52646	0.02734
2	0.67119	0.01759	0.00171
3	0.04383	0.07542	0.00263

Tabel 5.3 *Centroid* untuk C=4

Center	Rec	Freq	Mon
<i>Cluster</i>			
2	0.70326	0.01211	0.00089
1	0.02127	0.22226	0.00829
4	0.05317	0.04417	0.00142
3	0.00816	0.67099	0.03557

Tabel 5.4 *Centroid* untuk C=5

Center	Rec	Freq	Mon
<i>Cluster</i>			
3	0.38829	0.02696	0.00178
5	0.80928	0.00856	0.00066
2	0.01174	0.2481	0.00893
1	0.03805	0.04898	0.00123
4	0.00363	0.70601	0.0362

**Tabel 5.5 Centroid untuk C=6**

Center	Rec	Freq	Mon
Centroid			
6	0.41711	0.01956	0.0014
2	0.83603	0.00702	0.00044
4	0.01444	0.14676	0.00342
3	0.04584	0.03546	0.00097
1	0.00887	0.39107	0.01243
5	0.003	0.76783	0.04628

**Tabel 5.6 Centroid untuk C=7**

Center	Rec	Freq	Mon
Cluster			
7	0.33552	0.02407	0.00124
1	0.89097	0.00442	0.00035
6	0.01207	0.14943	0.0033
5	0.04026	0.03682	0.00091
2	0.00701	0.3972	0.01179
4	0.62952	0.01146	0.00069
3	0.0018	0.76563	0.04256

**Tabel 5.7 Centroid untuk C=8**

Center	Rec	Freq	Mon
Cluster			
7	0.35901	0.02265	0.00099

Center	Rec	Freq	Mon
Cluster			
6	0.89278	0.00395	0.00026
8	0.01871	0.10357	0.00192
5	0.04838	0.03023	0.00081
2	0.00643	0.46869	0.01122
1	0.00846	0.23095	0.00641
4	0.63648	0.01005	0.0005
3	0.00143	0.84654	0.05498

## 5.2. Validasi dengan Metode SSE

Setelah mendapatkan hasil *clustering* dari masing-masing jumlah *cluster* dari 3 sampai 8. Selanjutnya adalah melakukan validasi terhadap hasil *clustering* itu dengan menggunakan metode *Sum Square Error* (SSE). Pada proses validasi juga digunakan aplikasi R untuk membantu pengerjaan proses ini.

Proses ini akan memasukkan rumus yang sudah ada ke dalam aplikasi R yang sudah dijelaskan pada bab 2.2.7.2. Berikut *coding* dari aplikasi R nya. Berikut langkah-langkah pengerjaan validasi SSE.

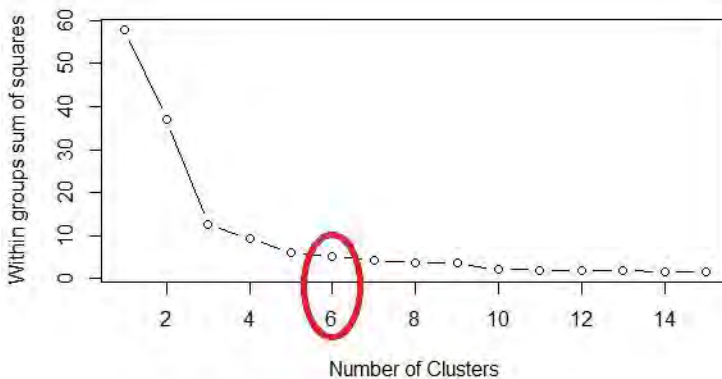
- Langkah pertama adalah menggunakan data set “dataku” yang telah diunggah pada proses *clustering*. Kemudian memasukkan rumus SSE ke R seperti *code* di bawah ini. Nilai SSE akan dicari mulai klaster 1 sampai 15.

```
wss <- (nrow(dataku)-1)*sum(apply(dataku,2,var))
> for (i in 1:15) wss[i] <- sum(cmeans(dataku
, centers=i)$withinss)
```

- Selanjutnya adalah membuat grafik dari hasil SSE dari klaster 1 sampai 15 dengan x axis jumlah klaster dan y axis nilai SSEnya.

```
plot(1:15, wss, type="b", xlab="Number of Clusters", ylab="within groups sum of squares")
```

Pada proses ini akan dipilih berapa jumlah *cluster* yang maksimal dengan mencari SSE dari masing-masing jumlah *cluster* dari 3 sampai 8. Aplikasi R mempermudah pengerjaan tugas akhir dengan menampilkan grafik dari hasil SSEnya. Pada hasil SSE semakin banyak *cluster* semakin baik, namun pada penjelasan di aplikasi R dipilih jumlah *cluster* dimana beda SSE dengan *cluster* kanannya hanya sedikit bedanya. Hasil dari proses ini dapat dilihat pada gambar 5.1.



**Gambar 5.1 Hasil Validasi SSE pada Aplikasi R**

Pada gambar di atas dapat dilihat bahwa SSE jumlah cluster 6 lebih stabil dibandingkan dengan SSE cluster lain ke setelahnya. Karena dari hasil SSE yang sudah dilakukan pada aplikasi R pada tugas akhir kali ini akan dipilih jumlah cluster 6.



### 5.3. Validasi dengan DBI

Setelah melalui proses validasi SSE, untuk membandingkan validasi maka akan dilanjutkan dengan menggunakan validasi DBI, untuk proses validasi DBI pada R, berikut *code* dari validasi dengan DBI.

- Langkah pertama adalah menggunakan data set “dataku” yang telah diunggah ke aplikasi R pada proses *clustering*. Kemudian mengaktifkan *package* yang dapat digunakan untuk mencari nilai DBI.

```
library(clusterSim)
```

- Langkah selanjutnya adalah menentukan metode untuk jaraknya dan menentukan berapa kluster minimal dan maksimal yang akan dicari.

```
md <- dist(dataku, method="euclidean")
> # nc - number_of_clusters
> min_nc=2
> max_nc=8
```

- Lalu membuat array untuk menyimpan nilai DBI

```
> res <- array(0, c(max_nc-min_nc+1, 2))
> res[,1] <- min_nc:max_nc
> clusters <- NULL
```

- Langkah selanjutnya adalah memasukkan rumus DBI ke dalam aplikasi R. Untuk mendapatkan nilai DBI mulai kluster minimal sampai maksimal yang telah ditentukan pada langkah sebelumnya.

```
> for (nc in min_nc:max_nc)
+ {
+   hc <- hclust(md, method="complete")
+   cl2 <- cutree(hc, k=nc)
+   res[nc-min_nc+1, 2] <- DB <- index.DB(d
+   ataku, cl2, centrotypes="centroids")$DB
+   clusters <- rbind(clusters, cl2)
+ }
```

- Langkah selanjutnya adalah mencetak hasil DBI pada masing-masing klaster dan mencetak juga di klaster mana nilai DBI paling minimal.

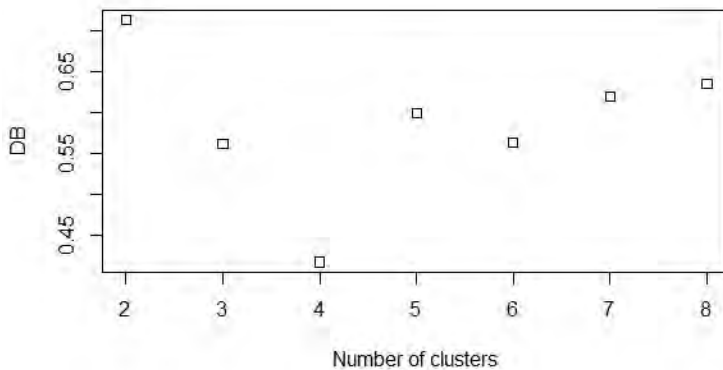
```
> print(paste("min DB for", (min_nc:max_nc)[which.min(res[,2])], "clusters=", min(res[,2])))
print("clustering for min DB")
print(clusters[which.min(res[,2]),])
write.table(res, file="DB_res.csv", sep=";", dec=".", row.names=TRUE, col.names=FALSE)
```

- Langkah terakhir adalah menampilkan hasil nilai DBI tersebut pada grafik sehingga mudah untuk memahami hasil DBI ini. Untuk x axisnya adalah jumlah klaster yang dicari dan y axis adalah nilai DBI.

```
> plot(res, type="p", pch=0, xlab="Number of clusters", ylab="DB", xaxt="n")
> axis(1, c(min_nc:max_nc))
```

Lalu *coding* di atas adalah untuk mencari DBI dari klaster 2 sampai 8 untuk dicari yang DBInya paling kecil. Lalu membuat grafik pada R sehingga mudah untuk dilihat perbedaannya.

Hasil dari DBI dapat dilihat pada gambar 5.2. Dari gambar dapat dilihat bahwa klaster 4 adalah klaster yang paling maksimal. Perbedaan hasil validasi ini diakibatkan karena memang DBI ini tidak kurang cocok untuk memvalidasi Fuzzy Clustering Method. Sementara SSE adalah validasi umum yang dapat digunakan pada semua metode *clustering*. Perbedaannya juga karena SSE menentukan klaster maksimalnya masih subyektif atau kira-kira.



**Gambar 5.2 Hasil Validasi DBI pada Aplikasi R**

#### **5.4. Pembuatan Visualisasi untuk Membantu Analisis**

Visualisasi dibentuk untuk membantu pengambil keputusan dari perusahaan untuk menganalisis pelanggan dari PT. XYZ. Pada tugas akhir ini akan dibuat dua visualisasi untuk membantu menganalisis pelanggan. Visualisasi yang pertama adalah visualisasi untuk menampilkan lokasi dari masing-masing pelanggan. Kemudian visualisasi yang kedua adalah visualisasi yang menampilkan transaksi yang dilakukan pelanggan per minggu baik dari banyaknya transaksi yang dilakukan dan besarnya skala transaksi yang dilakukan.

##### **5.4.1. Visualisasi Scatter Plot**

Pembuatan visualisasi ini akan menghasilkan *scatter plot* 3D (Bab 2.2.8.3) yang memperlihatkan persebaran masing-masing atribut dari recency, frequency, dan monetary. Visualisasi ini akan membantu analisis melakukan analisis RFM.

Untuk menghasilkan visualisasi tersebut, maka berikut langkah kerjanya:

- Langkah pertama adalah mengambil data hasil normalisasi RFM pada proses normalisasi dan menggabungkannya dengan hasil proses *clustering*. Kemudian dilakukan pemrograman pada aplikasi R. Langkah pertama pada aplikasi R adalah mengunggah data csv hasil *clustering* dan normalisasi ke aplikasi R.

```
scatterplot <- read.csv("scatterplot.csv",
header=T)
```

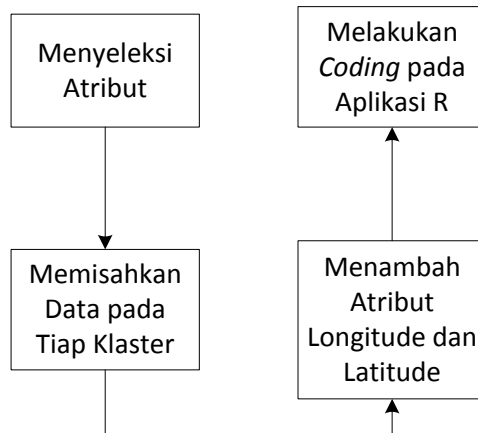
- Kemudian akan dilakukan pemberian warna yang berbeda untuk masing-masing klaster yang ada. Seperti warna merah untuk klaster 1, biru untuk klaster 2, hijau untuk klaster 3, hitam untuk klaster 4, klaster 5 berwarna merah muda, dan klaster 6 berwarna kuning.

```
dataku$pcolor[dataku$Cluster==1] <- "red"
dataku$pcolor[dataku$Cluster==2] <- "blue"
dataku$pcolor[dataku$Cluster==3] <- "green"
dataku$pcolor[dataku$Cluster==4] <- "black"
dataku$pcolor[dataku$Cluster==5] <- "pink"
dataku$pcolor[dataku$Cluster==6] <- "yellow"
```

- Langkah selanjutnya adalah membuat grafiknya dengan memanggil data yang sudah kita unggah ke aplikasi R dan menentukan x axis, y axis, dan z axis pada data yang kita unggah dan menentukan warna berdasarkan code yang telah dibuat pada langkah sebelumnya.

```
with(scatterplot, {
+   scatterplot3d(RecNorm,    # x axis
+                 FreqNorm,   # y axis
+                 MonNorm,    # z axis
+                 color = pcolor,
+                 main="3-D Scatterplot Example 1")
+ })
```

#### 5.4.2. Visualisasi dengan *googleVis*



**Gambar 5.3 Langkah Pembuatan Visualisasi dengan *Package GoogleVis***

Pembuatan visualisasi dengan *googleVis* (Bab 2.2.8.3) ini diharapkan menghasilkan visualisasi yang menunjukkan persebaran lokasi pelanggan pada google maps. Hal itu akan dapat dimanfaatkan oleh PT. XYZ untuk menentukan apakah pelanggan pada klaster tertentu itu jaraknya jauh atau dekat dari kota Semarang yang merupakan tempat dari PT. XYZ.

Langkah-langkah besar untuk membuat visualisasi ini dapat dilihat pada gambar 5.3. Untuk langkah-langkah yang lebih detail adalah sebagai berikut:

- Langkah pertama yang harus dilakukan dalam membuat visualisasi ini adalah menyeleksi atribut pada data yang ada dan hanya mengambil atribut yang diperlukan seperti KODE LANGGANAN dan ALAMAT LANGGANAN. Contoh dari hasil penyeleksian atribut dapat dilihat pada tabel 5.8.

**Tabel 5.8 Cuplikan Hasil Seleksi Atribut untuk Membuat Visualisasi Map**

Kode Langganan	Alamat
515049	XXXXX, SEMARANG
96033	XXXXX, SEMARANG.
86870	XXXXXX, PALEBON PEDURUNGAN SEMARANG
95783	XXXXXX, PLAMONGANSARI, PEDURUNGAN, SEMARANG
86886	XXXXX, SEMARANG
86725	XXXXX, SEMARANG
86620	XXXXX, SMG.
725451	XXXXXX
529177	XXXXX, SEMARANG
876484	XXXXXX, SEMARANG
56734	XXXXX, KALICARI, PEDURUNGAN, SMG
86879	XXXXX, KRANGGAN, SEMARANG TENGAH
58537	XXXXXX, NGEMPLAK SIMONGAN, SEMARANG
930012	XXXXX

- Kemudian mengumpulkannya pada setiap klaster yang telah dibentuk sebelumnya untuk mempermudah melakukan *coding* pada aplikasi R. Contoh hasil dari proses ini dapat dilihat pada tabel 5.9 yang merupakan kumpulan dari klaster 2.

**Tabel 5.9 Cuplikan Data Pengumpulan pada Klaster**

Nama Langganan	Alamat
9023591	XXXX

9023588	XXXXX
549913	XXXXX SEMARANG
9023587	XXXXXX
594814	XXXXX
86838	XXXXXX, SEMARANG.
86674	XXXXXX, SMG.
614454	XXXXXX
902355	XXXXX

- Kemudian dilanjutkan dengan menambahkan atribut *longitude* dan *latitude* karena fungsi *googleVis* pada aplikasi R hanya dapat membaca *longitude* dan *latitude* untuk menentukan lokasinya. Maka dari itu akan ditambah atribut ini dengan bantuan *googleMaps*. Contoh hasil dari proses ini ada pada tabel 5.10.

**Tabel 5.10 Cuplikan Hasil Penambahan Atribut *Longitude* dan *Latitude***

Nama Langganan	Alamat	Lokasi
9023591	XXXX	-7.XXX:110.XXXX
9023588	XXXXX	-6.XXXX:110.XXXX
549913	XXXXXX, SEMARANG	-7.XXXX:110.XXXX
9023587	XXXXXX	-7.XXXX:110.XXXX
594814	XXXXXXX	-7.XXXX:110.XXXX
86838	XXXXXX	-7.XXXX:110.XXX
86674	XXXXXX	-7.XXXX:109.XXXX
614454	XXXXXXX	-7.XXXX:110.XXXX
902355	XXXXX	-7.XXXX:110.XXXX

- -7.XXX itu menunjukkan *longitude* dan 110.XXX menunjukkan *latitude*. Langkah selanjutnya adalah pemrograman yang dilakukan pada aplikasi R. Pertama, data

csv yang ada *longitude latitude* nya diunggah ke aplikasi R. Kemudian mengaktifkan *package* yang dibutuhkan.

```
Cluster1 <- read.csv("LokasiCluster1.csv", header=T)
library("googlevis", lib.loc="C:/Program Files/R/R-3.1.2/library")
```

- Selanjutnya adalah membuat variabel yang digunakan untuk menampilkan map pada langkah selanjutnya. Pada *code* di bawah ini ditentukan data yang digunakan dan menentukan lokasinya berdasarkan atribut apa, pada tugas akhir ini ada pada atribut “Lokasi”. Dan mapnya juga dapat digerakkan, dapat di perbesar atau perkecil, dan juga menentukan tipe mapnya yaitu *terrain*.

```
Cluster1Map <- gvisMap(Cluster1, "Lokasi", options=list(showTip=TRUE, showLine=TRUE, enableScrollwheel=TRUE, mapType='terrain', useMapTypeControl=TRUE))
```

- Langkah selanjutnya adalah menampilkan variabel yang kita buat itu menjadi sebuah grafik atau visualisasi.

```
plot(Cluster1Map)
```

#### 5.4.3. Visualisasi dengan *HeatMap*

Pembuatan visualisasi ini akan menghasilkan *heatmap* (Bab 2.2.8.3) yang memperlihatkan *frequency* dan *monetary* tiap minggu pada setiap bulan. Hal ini akan dimanfaatkan PT. XYZ untuk mengetahui transaksi paling banyak dilakukan pada minggu ke berapa bulan ke berapa, begitu juga dengan pada minggu dan bulan apa pelanggan melakukan transaksi dalam jumlah yang besar. Dan juga pada minggu ke berapa pelanggan biasa melakukan transaksi.

Untuk menghasilkan visualisasi tersebut, maka berikut langkah kerjanya:

- Langkah pertama yang dilakukan adalah memisahkan *frequency* dan *monetary* per minggu pada tiap bulannya dan juga pada masing-masing klaster. Contoh hasil dari proses ini



dapat dilihat pada tabel 5.11. Tabel 5.11 tersebut kemudian disimpan dalam bentuk csv.

**Tabel 5.11 Cuplikan Data yang Digunakan untuk Membuat Visualisasi  
*HeatMap***

Bulan	Week1	Week2	Week3	Week4
Januari2013	434	527	403	440
Februari2013	449	480	399	460
Maret2013	486	382	422	331
April2013	473	462	449	364
Mei2013	493	363	437	357
Juni2013	335	361	400	344
Juli2013	363	482	438	464
Agustus2013	340	255	394	302
September2013	386	440	410	374
Oktober2013	393	314	417	437
November2013	357	318	280	420
Desember2013	339	351	445	338
Januari2014	300	339	389	314
Februari2014	361	377	356	403
Maret2014	368	318	399	323
April2014	434	422	338	375
Mei2014	370	377	377	303
Juni2014	347	383	409	265
Juli2014	374	312	460	139
Agustus2014	346	308	318	290
September14	338	324	324	237
Oktober2014	285	331	296	284
November14	279	265	310	176

- Setelah itu dilakukan pemrograman dengan aplikasi R. Langkah pertama pada aplikasi R adalah mengunggah csv yang telah dibuat pada langkah sebelum ini. Kemudian mengurutkan data berdasarkan week1 dan menjadikan bulan sebagai variabel untuk keperluan langkah selanjutnya.

```
Cluster1 <- read.csv("Clust1.csv", header=T)
Cluster1$Bulan <- with(Cluster1, reorder(Bulan, week1))
```

- Langkah selanjutnya adalah mengaktifkan *package* ggplot2, reshape2, plyr, dan scales untuk keperluan langkah selanjutnya.

```
library("ggplot2", lib.loc="C:/Program Files/R/R-3.1.2/library")
library("reshape2", lib.loc="C:/Program Files/R/R-3.1.2/library")
library("plyr", lib.loc="C:/Program Files/R/R-3.1.2/library")
library("scales", lib.loc="C:/Program Files/R/R-3.1.2/library")
```

- Langkah selanjutnya adalah melakukan rescale atau pemberian ulang skala. Namun sebelum langkah tersebut data harus dimelt terlebih dahulu.

```
Cluster1.m <- melt(Cluster1)
Cluster1.m <- dply(Cluster1.m, .(variable),
transform, rescale = rescale(value))
```

- Langkah terakhir adalah membuat grafiknya dengan memakai variabel yang telah ditentukan diawal tadi dan skalanya menggunakan skala baru yang sudah dibuat. Dan juga membuat gradasi warna dimana yang paling rendah akan berwarna putih dan yang paling banyak akan berwarna biru.

```
(p <- ggplot(Cluster1.m, aes(variable, Bulan))
+ geom_tile(aes(fill = rescale), colour = "white")
+ scale_fill_gradient(low = "white", high = "steelblue"))
```

## **BAB VI**

### **HASIL DAN PEMBAHASAN**

Pada bab ini akan dijelaskan bagaimana karakteristik dari setiap *cluster* berdasarkan analisis dari attribute yang digunakan untuk proses *clustering*. Bab ini juga berisikan tentang analisis yang diambil dari kebiasaan dan demografi dari setiap *cluster*.

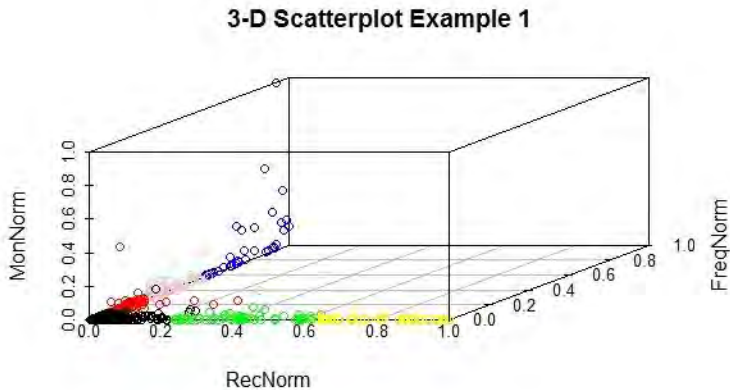
#### **6.1. Analisis Hasil *Clustering***

Setelah menemukan hasil *cluster* maksimal, langkah selanjutnya adalah melakukan analisis terhadap setiap *cluster* untuk menemukan karakteristik pelanggan setiap *cluster*. Analisis ini akan dilakukan berdasarkan dasar teori dengan bab 2.2. Analisis ini akan dibagi berdasarkan nilai RFM, analisis geografis, *behavior*, dan tipe pelanggan. Analisis geografis dilakukan dengan melihat letak dari setiap pelanggan yang ada di setiap *cluster*, analisis tipe pelanggan dilakukan dengan memisahkan pelanggan menjadi apotek, rumah sakit dan lain-lain. Kemudian analisis *behavior* dilakukan dengan melihat kebiasaan transaksi dan besaran transaksi yang dilakukan oleh pelanggan pada setiap *cluster*. Beberapa analisis tersebut diharapkan akan membentuk karakteristik dari setiap pelanggan dengan *detail*.

##### **6.1.1. Analisis Antar Atribut RFM**

Analisis hasil RFM ini juga dapat digunakan untuk menentukan karakteristik dari setiap klaster pelanggan perusahaan dan juga membantu analisis lainnya. Untuk membantu analisis ini akan ditampilkan *scatter plot* dari semua pelanggan dengan dipisahkan warna untuk setiap klasternya. Pemilihan *scatter plot* telah dijelaskan pada bab 2.2.8.3. Analisis juga akan dilakukan dengan membandingkan antara satu atribut RFM dengan atribut lainnya. Dengan melihat *scatter plot* diharapkan dapat membantu analisis ini menjadi lebih jelas. Pada gambar 6.x dapat dilihat bahwa x axis adalah recency, y axis adalah monetary, dan z axis adalah frequency. Untuk warnanya, klaster pertama berwarna merah; klaster 2 berwarna biru, klaster 3 berwarna hijau; klaster 4

berwarna hitam; klaster 5 berwarna merah muda; dan klaster 6 berwarna kuning.



**Gambar 6.1 Scatter plot 3D (RFM)**

Untuk analisis antar atribut *recency* dan *frequency* bisa dilihat pada gambar 12 dan hasilnya dapat dilihat pada tabel 6.13.

**Tabel 6.1 analisis *cluster* antara atribut *recency*&*frequency***

Cluster	Hasil Analisis
1 (Merah)	<i>Cluster 1</i> memiliki rata-rata <i>recency</i> 91 hari yang memiliki selisih <b>jauh lebih sedikit</b> dengan rata-rata <i>recency</i> semua pelanggan. Sementara itu nilai <i>frequency</i> yaitu 94 kali dan juga memiliki selisih <b>lebih besar sedikit</b> daripada rata-rata semua pelanggan. Dapat disimpulkan bahwa pelanggan ini adalah pelanggan yang lumayan sering melakukan transaksi hingga akhir-akhir ini.
2 (Biru)	Nilai <i>recency</i> dari <i>cluster 2</i> rata-rata adalah 80 hari yang berarti nilai <i>recency</i> <b>cukup baik</b> karena

	selisihnya jauh lebih sedikit daripada rata-rata semua pelanggan. Sementara <i>frequency</i> dari <i>cluster</i> 2 adalah 442 kali, yang berarti <i>cluster</i> ini <b>sangat sering</b> melakukan transaksi karena jauh dari rata-rata semua pelanggan. Hal ini menunjukkan bahwa pelanggan pada <i>cluster</i> ini sangat sering melakukan transaksi dan sangat stabil sampai akhir dilakukan analisis ini.
3 (Hijau)	<i>Cluster</i> 3 memiliki nilai <i>recency</i> rata-rata yang <b>sangat besar</b> yang berarti <b>sangat buruk</b> karena nilai <i>recency</i> pada <i>cluster</i> ini adalah 363 hari. Selisihnya sangat besar sekali dari rata-rata pelanggan keseluruhan. Untuk nilai <i>frequency</i> dari <i>cluster</i> ini adalah 14 kali, yang artinya berselisih jauh <b>lebih kecil</b> dari rata-rata semua pelanggan. Dari hasil tersebut dapat diartikan bahwa pelanggan pada <i>cluster</i> ini melakukan transaksi sangat sedikit dan juga transaksi itu sudah sangat lama dilakukannya.
4 (Hitam)	<i>Cluster</i> 4 memiliki nilai <i>recency</i> rata-rata 119 hari yang memiliki selisih <b>lebih sedikit</b> daripada rata-rata semua pelanggan. Untuk nilai <i>frequency</i> dari <i>cluster</i> ini yaitu 21 kali yang selisihnya <b>jauh lebih sedikit</b> daripada rata-rata semua pelanggan. Hal ini menunjukkan bahwa pelanggan pada <i>cluster</i> ini adalah pelanggan baru yang memiliki nilai rata-rata <i>recency</i> dan <i>frequency</i> sedikit.
5 (Merah muda)	Nilai <i>recency</i> dari <i>cluster</i> ini rata-rata adalah 85 hari yang selisihnya <b>jauh lebih sedikit</b> dari rata-rata semua pelanggan. Untuk nilai <i>frequency</i> dari <i>cluster</i> ini rata-rata adalah 226 kali, yang selisihnya <b>jauh lebih besar</b> dari rata-rata semua pelanggan. Hal ini menunjukkan bahwa pelanggan pada <i>cluster</i> ini adalah sangat sering melakukan transaksi dan sangat stabil sampai akhir-akhir ini.
6 (Kuning)	<i>Cluster</i> 6 ini memiliki nilai <i>recency</i> rata-rata 628 hari yang bedanya <b>sangat jauh</b> dari rata-rata semua

	pelanggan, yang artinya nilai <i>recency</i> <b>sangat buruk</b> . Untuk nilai <i>frequency</i> nya juga <b>sangat sedikit</b> yaitu rata-rata 5. Hal ini menunjukkan bahwa rata-rata pelanggan dari <i>cluster</i> ini adalah pelanggan yang sangat jarang melakukan transaksi dan sudah lama juga dalam melakukan transaksi.
--	--

Selanjutnya analisis akan dilanjutkan dengan melihat antara nilai *recency* dan *monetary*. Hasil analisis bisa dilihat pada tabel 6.14.

**Tabel 6.2 analisis cluster antara atribut *recency* & *monetary***

Cluster	Hasil Analisis
1 (Merah)	<i>Cluster 1</i> memiliki rata-rata <i>recency</i> 91 hari yang memiliki selisih <b>jauh lebih sedikit</b> dengan rata-rata <i>recency</i> semua pelanggan. Untuk rata-rata nilai <i>monetary</i> dari <i>cluster</i> ini adalah Rp 15.539.702 yang berselisih <b>lebih sedikit</b> daripada rata-rata semua pelanggan. Jadi pelanggan pada <i>cluster</i> ini adalah pelanggan yang transaksinya akhir-akhir ini namun dalam skala kecil.
2 (Biru)	Nilai <i>recency</i> dari <i>cluster 2</i> rata-rata adalah 80 hari yang berarti nilai <i>recency</i> <b>cukup baik</b> karena selisihnya jauh lebih sedikit daripada rata-rata semua pelanggan. Untuk nilai <i>monetary</i> dari <i>cluster</i> ini rata-rata adalah Rp 1.933.658.117 yang memiliki nilai <b>sangat besar</b> jika dibandingkan dengan rata-rata semua pelanggan. Hal ini menunjukkan bahwa pelanggan pada <i>cluster</i> ini adalah pelanggan yang sangat banyak jika menghabiskan uang dan juga transaksinya dilakukan akhir-akhir ini.
3 (Hijau)	<i>Cluster 3</i> memiliki nilai <i>recency</i> rata-rata yang <b>sangat besar</b> yang berarti <b>sangat buruk</b> karena nilai <i>recency</i> pada <i>cluster</i> ini adalah 363 hari. Selisihnya sangat besar sekali dari rata-rata pelanggan keseluruhan.

	Untuk nilai <i>monetary</i> dari <i>cluster</i> ini rata-rata adalah Rp 10.408.746. Yang artinya nilai <i>monetary</i> <b>kecil</b> . Hal ini menunjukkan bahwa pelanggan pada <i>cluster</i> ini pelanggan yang sudah jarang melakukan transaksi dan transaksinya juga dalam skala kecil.
4 (Hitam)	<i>Cluster</i> 4 memiliki nilai <i>recency</i> rata-rata 119 hari yang memiliki selisih <b>lebih sedikit</b> daripada rata-rata semua pelanggan. Untuk nilai <i>monetary</i> dari <i>cluster</i> ini rata-rata adalah Rp 26.865.483 yang selisihnya juga lumayan <b>jauh lebih sedikit</b> daripada rata-rata semua pelanggan. Berdasarkan hasil ini, pelanggan pada <i>cluster</i> ini adalah pelanggan yang baru saja melakukan transaksi, namun masih sedikit skala transaksi yang dilakukan.
5 (Merah muda)	Nilai <i>recency</i> dari <i>cluster</i> ini rata-rata adalah 85 hari yang selisihnya <b>jauh lebih sedikit</b> dari rata-rata semua pelanggan. Untuk rata-rata nilai <i>monetary</i> dari <i>cluster</i> ini adalah Rp 368.189.758 yang selisihnya <b>cukup jauh lebih besar</b> daripada rata-rata semua pelanggan. Hal ini menunjukkan bahwa pelanggan pada <i>cluster</i> ini akhir-akhir ini melakukan transaksi dan transaksinya juga dilakukan dalam skala yang lumayan besar.
6 (Kuning)	<i>Cluster</i> 6 ini memiliki nilai <i>recency</i> rata-rata 628 hari yang bedanya <b>sangat jauh</b> dari rata-rata semua pelanggan, yang artinya nilai <i>recency</i> <b>sangat buruk</b> . Untuk nilai rata-rata <i>monetary</i> dari <i>cluster</i> 6 ini adalah Rp 2.519.902 yang berarti <b>sangat kecil</b> . Dari hasil tersebut maka pelanggan pada <i>cluster</i> ini adalah pelanggan yang sudah jarang melakukan transaksi dan transaksinya juga dalam skala yang sangat kecil.

Untuk selanjutnya akan dianalisis hasil *cluster* berdasarkan atribut *frequency* dan *monetary*. Hasil analisisnya bisa dilihat pada tabel 6.15.

Tabel 6.3 analisis *cluster* antara atribut *frequency* & *monetary*

Cluster	Hasil Analisis
1 (Merah)	Nilai <i>frequency</i> yaitu 94 kali dan juga memiliki selisih <b>lebih besar sedikit</b> daripada rata-rata semua pelanggan. Untuk rata-rata nilai <i>monetary</i> dari <i>cluster</i> ini adalah Rp 15.539.702 yang berselisih <b>lebih sedikit</b> daripada rata-rata semua pelanggan. Berdasarkan hasil tersebut berarti pelanggan pada <i>cluster</i> ini sering melakukan transaksi namun transaksinya dalam skala kecil.
2 (Biru)	Nilai <i>frequency</i> dari <i>cluster</i> 2 adalah 442 kali, yang berarti <i>cluster</i> ini <b>sangat sering</b> melakukan transaksi karena jauh dari rata-rata semua pelanggan. Untuk nilai <i>monetary</i> dari <i>cluster</i> ini rata-rata adalah Rp 1.933.658.117 yang memiliki nilai <b>sangat besar</b> jika dibandingkan dengan rata-rata semua pelanggan. Hal ini menunjukkan bahwa pelanggan pada <i>cluster</i> ini sangat sering melakukan transaksi dan transaksinya juga dalam jumlah yang sangat besar.
3 (Hijau)	Untuk nilai <i>frequency</i> dari <i>cluster</i> ini adalah 14 kali, yang artinya berselisih jauh <b>lebih kecil</b> dari rata-rata semua pelanggan. Untuk nilai <i>monetary</i> dari <i>cluster</i> ini rata-rata adalah Rp 10.408.746. Yang artinya nilai <i>monetary</i> <b>kecil</b> . Yang berarti bahwa pelanggan pada <i>cluster</i> ini jarang melakukan transaksi dan transaksinya dalam jumlah kecil.
4 (Hitam)	Untuk nilai <i>frequency</i> dari <i>cluster</i> ini yaitu 21 kali yang selisihnya <b>jauh lebih sedikit</b> daripada rata-rata semua pelanggan. Untuk nilai <i>monetary</i> dari <i>cluster</i> ini rata-rata adalah Rp 26.865.483 yang selisihnya juga lumayan <b>jauh lebih sedikit</b> daripada rata-rata semua pelanggan. Hal ini menunjukkan bahwa pelanggan dari <i>cluster</i> ini jarang melakukan transaksi dan transaksinya juga masih belum terlalu banyak.



5 (Merah muda)	Untuk nilai <i>frequency</i> dari <i>cluster</i> ini rata-rata adalah 226 kali, yang selisihnya <b>jauh lebih besar</b> dari rata-rata semua pelanggan. Untuk rata-rata nilai <i>monetary</i> dari <i>cluster</i> ini adalah Rp 368.189.758 yang selisihnya <b>cukup jauh lebih besar</b> daripada rata-rata semua pelanggan. Hal ini menunjukkan bahwa pelanggan pada <i>cluster</i> ini cukup sering melakukan transaksi dan transaksinya juga cukup banyak skalanya.
6 (Kuning)	Untuk nilai <i>frequency</i> nya juga <b>sangat sedikit</b> yaitu rata-rata 5. Untuk nilai rata-rata <i>monetary</i> dari <i>cluster</i> 6 ini adalah Rp 2.519.902 yang berarti <b>sangat kecil</b> . Karena nilai <i>frequency</i> dan <i>monetary</i> yang sangat kecil maka pelanggan pada <i>cluster</i> ini sangat jarang melakukan transaksi dan sangat kecil jika melakukan transaksi.

### 6.1.2. Hasil Analisis Geografi

Untuk membantu analisis geografi yang pentingnya telah dijelaskan pada bab 2.2.2.1. maka pada tugas akhir ini akan dibuatkan visualisasi untuk memperlihatkan persebaran lokasi pelanggan pada setiap klaster. Analisis geografi ini menggunakan *package* googleVis pada aplikasi R. Cara membuatnya sudah dicantumkan pada bab 2.2.8.1. Berikut adalah laporan hasil analisis geografi.

Menurut wawancara dengan PT. XYZ analisis geografi (2.2.2.1) ini akan membantu mencari bagaimana persebaran pelanggan pada setiap klaster sehingga dapat dimanfaatkan oleh PT. XYZ untuk menentukan cara promosi dan pengiriman barang yang akan dilakukan.

Untuk hasil analisis dari *cluster* 1 dapat dilihat pada gambar 6.1.



**Gambar 6.2 Lokasi Pelanggan Klaster 1**

Berdasarkan gambar 6.1 di atas maka bisa dilihat bahwa pelanggan pada klaster 1 persebarannya cukup menumpuk pada kota Semarang, hanya kurang dari setengah yang lokasi pelanggannya di luar kota Semarang dan hanya ada satu pelanggan yang tempatnya sangat jauh dari kota Semarang. Data dari gambar 6.1 di atas didapatkan dari 107 pelanggan yang berada pada klaster 1. Maka dari itu dapat disimpulkan bahwa pelanggan pada klaster ini lokasinya lebih banyak yang berada di dalam kota Semarang namun tidak sedikit juga yang berada cukup jauh dari kota Semarang.

Karena lokasi pelanggan pada klaster ini masih berada di kota Semarang maka tidak diperlukan adanya batasan minimal pembelian karena biaya pengiriman lebih sedikit dan untuk melakukan promosi juga mudah dijangkau. Namun yang bermasalah adalah beberapa pelanggan yang jaraknya lumayan jauh dan juga jika dilihat dari analisis RFMnya adalah pelanggan ini pelanggan yang suka melakukan transaksi namun tidak dalam skala banyak. Cara untuk mengatasi pengirimannya adalah mengumpulkan transaksi pelanggan yang lumayan jauh kemudian mengirimnya secara bersamaan.

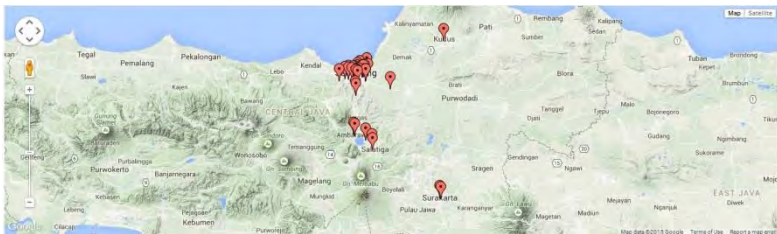
Untuk hasil analisis pelanggan pada klaster 2 bisa dilihat pada gambar 6.2.

Pada gambar 6.2 tersebut dapat dilihat bahwa pelanggan yang berada pada klaster 2 ini berada pada daerah kota Semarang, hanya beberapa pelanggan yang berasal dari luar kota Semarang.

Data gambar 6.2 di atas didapatkan dari 38 pelanggan yang berada pada kluster 2. Dari analisis di atas dapat disimpulkan bahwa lokasi pelanggan pada kluster 2 ini adalah pelanggan yang berlokasi lebih banyak di kota Semarang namun ada beberapa pelanggan yang jaraknya cukup jauh.

Sama dengan kluster sebelumnya, karena lokasinya yang sangat atau di kota Semarang maka tidak diperlukan adanya batasan minimal pembelian produk dan promosi juga akan mudah dilakukan karena masih mudah dijangkau. Untuk yang jaraknya agak jauh, maka tidak ada masalah karena transaksi yang dilakukan pelanggan pada kluster ini selalu dalam jumlah besar jika dilihat pada analisis RFMnya.

Selanjutnya adalah analisis pelanggan pada kluster 3 yang ditampilkan pada gambar 6.3.



**Gambar 6.3 Lokasi Pelanggan Kluster 2**

Pada gambar 6.3 terlihat bahwa pelanggan pada kluster 3 ini lokasinya memang lebih banyak yang berada dalam kota Semarang, namun ada beberapa pelanggan yang memang jarak dari kota Semarang jauh. Data ini berasal dari total 58 pelanggan yang berada pada kluster 3. Hal tersebut menandakan bahwa pelanggan pada kluster 3 banyak yang dari dalam kota Semarang, namun tidak sedikit juga yang jaraknya jauh dari kota Semarang.

Dengan melihat analisis RFM yang sudah dilakukan pelanggan ini sudah jarang melakukan transaksi, perlu adanya upaya peningkatan loyalitas. Hal tersebut akan mudah dilakukan pada pelanggan yang berada di kota Semarang. Namun untuk

pelanggan yang jauh dari kota Semarang akan susah untuk menaikkan loyalitasnya. Untuk masalah pengirimannya PT. XYZ akan membuat pelanggan klaster ini mengambil sendiri produk yang dibelinya. PT. XYZ mengakui kemungkinan lokasi yang agak jauh inilah yang membuat PT. XYZ susah melakukan transaksi dan mengakibatkan nilai RFM yang rendah.



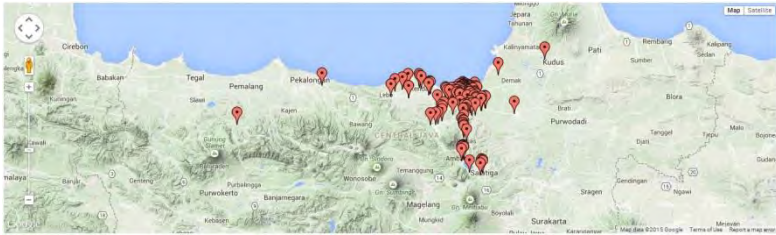
**Gambar 6.4 Lokasi Pelanggan Klaster 3**

Untuk pelanggan yang berada pada klaster 4 hasil analisis geografisnya bisa dilihat pada gambar 6.4.

Berdasarkan pada gambar 6.4 dapat dilihat bahwa pelanggan yang berada pada klaster 4 ini menumpuk pada kota Semarang. Hal ini diakibatkan karena pelanggan pada klaster ini berjumlah 164 pelanggan. Namun jika dilihat lagi bahwa pelanggan pada klaster ini juga banyak yang berasal dari luar kota Semarang dan jaraknya dari kota Semarang juga cukup jauh.

Dengan melihat analisis RFMnya pelanggan pada klaster ini masih berkesempatan untuk menjadi pelanggan yang loyalitasnya lebih baik lagi karena akhir-akhir ini masih melakukan transaksi. Oleh karena itu, untuk pelanggan yang jaraknya lumayan jauh dari kota Semarang diharapkan diberi promosi agar biasa melakukan transaksi dalam jumlah yang banyak.

Untuk pelanggan yang berada pada klaster 5 hasil analisis geografisnya bisa dilihat pada gambar 6.5.



**Gambar 6.5 Lokasi Pelanggan Klaster 4**

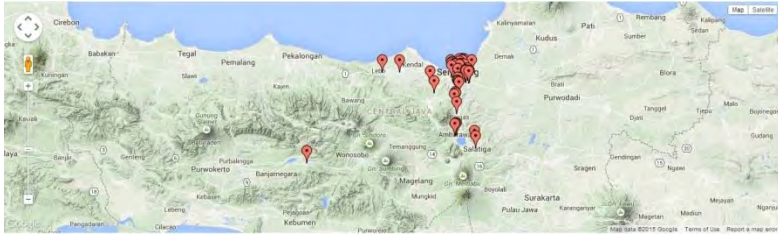


**Gambar 6.6 Lokasi Pelanggan Klaster 5**

Berdasarkan pada gambar 6.5 dapat dilihat bahwa pelanggan pada klaster 5 ini mayoritas berada pada kota Semarang, hanya ada sekitar 5 pelanggan yang berada di luar kota Semarang. Dari kelima pelanggan itu hanya 1 pelanggan yang jaraknya cukup jauh dari kota Semarang. Data ini diambil dari 59 pelanggan yang berada pada klaster 5.

Jika dilihat dari analisis RFMnya maka pelanggan ini sangat loyal dan jaraknya juga di sekitar kota Semarang saja. Jadi tidak ada masalah untuk pengiriman.

Untuk pelanggan yang berada pada klaster 6 hasil analisis geografisnya bisa dilihat pada gambar 6.6.



**Gambar 6.7 Lokasi Pelanggan Klaster 6**

Berdasarkan pada gambar 6.6 dapat dilihat bahwa pelanggan pada klaster 6 ini cukup menyebar. Jika dilihat memang sedikit lebih banyak pelanggan yang ada di kota Semarang, namun banyak pelanggan yang juga berada di luar kota Semarang yang jaraknya juga jauh dari kota Semarang. Data ini diambil dari 49 data pelanggan yang ada pada klaster 6.

Jika dilihat dari analisis RFMnya, klaster ini hampir sama dengan klaster 3 namun klaster ini lebih buruk. Maka cara untuk memperlakukan pelanggan pada klaster ini sama dengan pelanggan klaster 3 untuk promosi dan pengirimannya.

Kesimpulan untuk analisis geografi dapat dilihat pada tabel 6.4.

**Tabel 6.4 Kesimpulan Hasil Analisis Geografi**

Klaster	RFM	Analisis Geografi	Pengiriman Pesanan	Promosi
1	Recency Bagus, Frequency Bagus, dan Monetary Jelek	Agak tersebar di luar kota	Menggabungkan pesanan dari pelanggan karena sering melakukan transaksi namun dalam skala kecil	Mudah dilakukan karena sering melakukan transaksi dan promosinya berfokus

Klaster	RFM	Analisis Geografi	Pengiriman Pesanan	Promosi
				pada peningkatan monetary
2	Recency Bagus, Frequency Bagus, dan Monetary Bagus	Agak tersebar di luar kota	Tidak bermasalah karena transaksi dilakukan dalam jumlah besar	Mudah dilakukan karena sering melakukan transaksi
3	Recency Jelek, Frequency Jelek, dan Monetary Jelek	Terkumpul dalam kota	Menggabungkan pesanan dengan pelanggan klaster lain yang berada pada satu kota	Mudah dilakukan karena masih berada pada dalam kota
4	Recency Bagus, Frequency Jelek, dan Monetary Jelek	Terkumpul dalam kota	Menggabungkan pesanan dengan pelanggan klaster lain yang berada pada satu kota	Mudah dilakukan karena masih berada pada dalam kota
5	Recency Bagus, Frequency Bagus, dan Monetary Bagus	Terkumpul dalam kota	Tidak bermasalah karena transaksi dilakukan dalam jumlah besar	Mudah dilakukan karena sering melakukan transaksi dan promosinya

Klaster	RFM	Analisis Geografi	Pengiriman Pesanan	Promosi
				berfokus pada peningkatan monetary
6	Recency Jelek, Frequency Jelek, dan Monetary Jelek	Tersebar di luar kota	Menggabungkan pesanan dengan pelanggan klaster lain yang berada pada satu kota	Promosi dilakukan saat pelanggan melakukan transaksi

### 6.1.3. Hasil Analisis Tipe Pelanggan

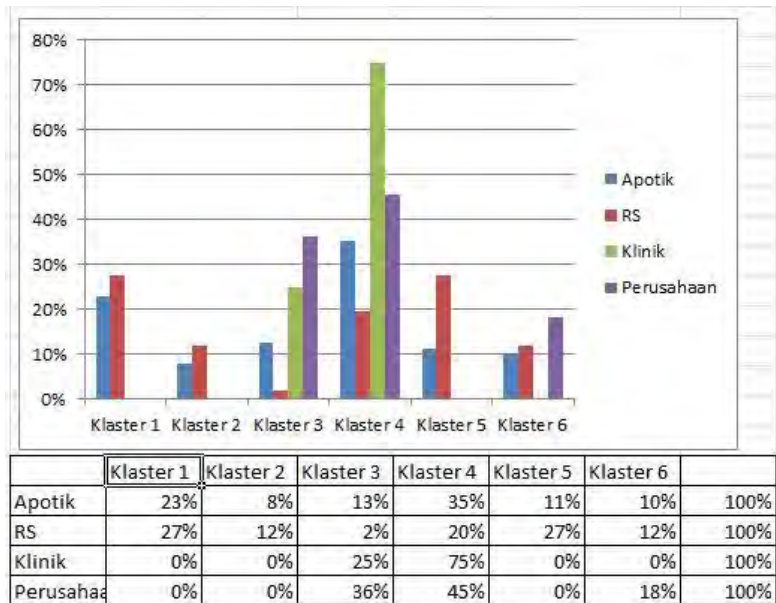
Berdasarkan wawancara dengan PT. XYZ analisis tipe pelanggan ini akan mencari secara kualitatif bagaimana persebaran tipe pelanggan pada setiap klaster. Menurut PT. XYZ selama ini mereka hanya mementingkan dua tipe pelanggan yaitu apotik dan rumah sakit karena jumlahnya sangat banyak.

Tetapi tidak semua pelanggan yang berjumlah banyak itu membawa keuntungan pada PT. XYZ, oleh karena itu akan dilakukan perbandingan antara kualitatif dan kuantitatif dari persebaran tipe pelanggan tersebut.

Analisis tipe pelanggan ini penting untuk memahami pelanggan, hal itu telah dicantumkan pada bab 2.2.2.2. Perusahaan ini memiliki 4 tipe pelanggan yaitu apotik (AP), rumah sakit (RS), klinik, dan perusahaan (PT). Analisis akan dilakukan dengan membandingkan 4 tipe pelanggan yang setipe pada klaster yang satu dengan yang lain. Jumlah masing-masing tipe yang akan dibandingkan adalah 405 AP, 51 RS, 8 klinik, dan 11 PT. Dengan jumlah tersebut setiap klaster pasti didominasi oleh apotik untuk jumlah paling banyak, namun analisis ini membandingkan



bagaimana setiap tipe tersebut persebarannya pada setiap kluster. Untuk membantu analisis maka akan dibuat diagram batang dengan menggunakan Microsoft excel. Gambar dari diagram batang yang telah dibuat ditampilkan pada gambar 6.7.



**Gambar 6.8 Hasil analisis demografi**

Dengan melihat gambar 6.7 maka dapat diambil kesimpulan bahwa:

- Tipe pelanggan apotik paling banyak menempati kluster 4, kemudian pada kluster 1, dan sisanya tersebar di kluster sisa. Hal ini diakibatkan tipe pelanggan apotik yang sangat banyak di PT. XYZ. Dan juga semakin banyaknya usaha apotik di daerah Jawa Tengah khususnya kota Semarang, maka dari itu tipe pelanggan apotik ini banyak di kluster 4 dan 1 yang sering melakukan transaksi akhir-akhir ini. Namun dalam skala yang tidak banyak.

- Tipe pelanggan rumah sakit ini paling banyak di klaster 5 dan klaster 1. Hal ini menunjukkan bahwa tipe pelanggan rumah sakit ini sangat sering dalam melakukan transaksi jika dilihat dari analisis RFM. Hal itu diakibatkan rumah sakit mempunyai dokter yang menyarankan membeli obat kepada pasien di apotik rumah sakit tersebut.
- Tipe pelanggan klinik hanya berada pada klaster 3 dan klaster 4, dan menempati terbanyak pada klaster 4. Hal ini menunjukkan bahwa semakin menjamurnya bisnis klinik yang ada di kota Semarang dan sekitarnya, dan obat yang dibeli hanya digunakan atau diberikan oleh bidan dari klinik tersebut secara langsung kepada pasien.
- Tipe pelanggan perusahaan/yayasan ini berada pada klaster 3, 4, dan 6. Ketiganya memiliki nilai RFM yang cukup buruk. Hal itu diakibatkan bahwa perusahaan/yayasan ini hanya melakukan transaksi jika butuh saja untuk anggota perusahaan atau yayasan mereka sendiri.

Selain melakukan analisis seperti di atas, juga dilakukan analisis dengan membandingkan kuantitas dan kualitas tipe pelanggan.

**Tabel 6.5 Jumlah Tipe Pelanggan Apotik dan Rumah Sakit pada Setiap Klaster**

Tipe Pelanggan: Apotik		Tipe Pelanggan: Rumah Sakit	
Klaster	Jumlah	Klaster	Jumlah
Klaster 4	35%	Klaster 5	27%
Klaster 1	23%	Klaster 1	27%
Klaster 3	13%	Klaster 4	20%
Klaster 5	11%	Klaster 2	12%
Klaster 6	10%	Klaster 6	12%
Klaster 2	8%	Klaster 3	2%

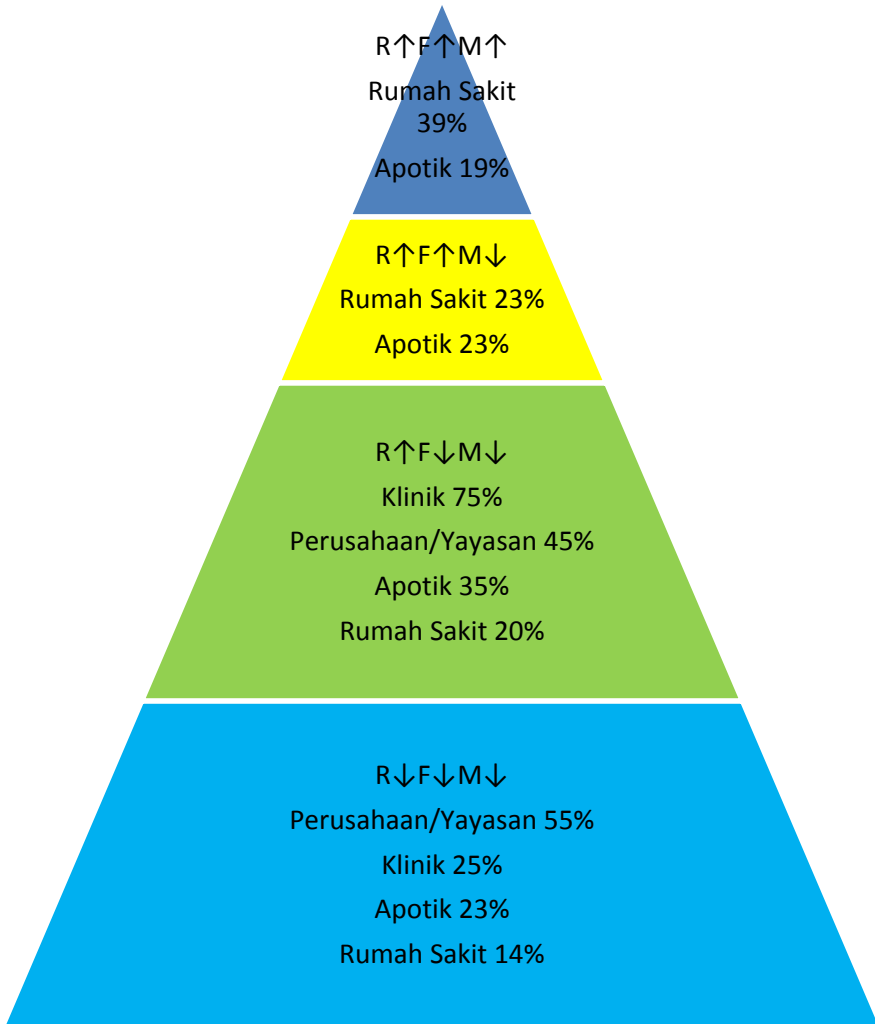
**Tabel 6.6 Jumlah Tipe Pelanggan Klinik dan Perusahaan/Yayasan pada Setiap Klaster**

Tipe Pelanggan: Klinik		Tipe Pelanggan: Perusahaan/Yayasan	
Klaster	Jumlah	Klaster	Jumlah
Klaster 4	75%	Klaster 4	45%
Klaster 3	25%	Klaster 3	36%
Klaster 1	0%	Klaster 6	18%
Klaster 2	0%	Klaster 2	0%
Klaster 6	0%	Klaster 1	0%
Klaster 3	0%	Klaster 5	0%

Pada tabel 6.5 dan 6.6 adalah jumlah masing-masing tipe pelanggan pada setiap klaster dengan urutan jumlah. Namun belum tentu semakin banyak pelanggan semakin baik, maka dari itu akan dihubungkan dengan analisis berdasarkan kualitas pelanggan yang hasilnya ada pada tabel 6.7.

**Tabel 6.7 Analisis Berdasarkan Kualitas Pelanggan (Pareto)**

Klaster	RFM
Klaster 2	R↑F↑M↑
Klaster 5	R↑F↑M↑
Klaster 1	R↑F↑M↓
Klaster 4	R↑F↓M↓
Klaster 3	R↓F↓M↓
Klaster 6	R↓F↓M↓



**Gambar 6.9 Urutan Tipe Pelanggan Berdasarkan Indeks RFM dari Terbaik Sampai Terburuk**

Berdasarkan tabel 6.5, 6.6, dan 6.7; dan gambar 6.9 dapat disimpulkan jika pada tipe pelanggan apotik ada 23% pelanggan yang RFMnya sangat buruk, maka PT. XYZ tidak perlu terlalu memperhatikan apotik yang termasuk ke dalam klaster 3 dan 6 yang jumlahnya 23% tersebut. PT. XYZ harus segera memperbaiki hubungan pelanggan karena apotik yang diharapkan menjadi pelanggan yang loyal masih ada 19%.

Untuk tipe pelanggan rumah sakit ada sekitar 14% dari pelanggan yang RFMnya sangat buruk dan tidak perlu diperhatikan oleh PT. XYZ. Harapan PT. XYZ untuk rumah sakit yang menjadi pelanggan loyal memenuhi harapan karena sebesar 39% pelanggan mempunyai RFM yang sangat baik.

Untuk tipe pelanggan klinik dan perusahaan atau yayasan tidak bisa diharapkan karena untuk klinik total semua pelanggannya mempunyai nilai RFM yang sangat buruk. Untuk perusahaan masih ada 45% pelanggan yang masih bisa ditingkatkan loyalitasnya.

#### **6.1.4. Hasil Analisis Behavior**

Berdasarkan wawancara dengan PT. XYZ, analisis *behavior* penting untuk memahami pelanggan dari perusahaan. Hal itu sudah dibahas pada bab 2.2.2.3. Manfaat dari laporan ini antara lain:

- Mencari hubungan pesanan dengan wabah / musim penyakit tertentu
- Mencari hubungan pesanan dengan apa yang terjadi dengan PT. XYZ cabang lainnya.
- Mencari hubungan transaksi dengan keputusan pemerintah.

Pemanfaatan analisis ini membutuhkan data yang sifatnya rahasia dan membutuhkan waktu untuk mengolahnya. Sehingga pada analisis ini lebih bersifat deskriptif membantu menjelaskan perilaku pelanggan yang telah terjadi.

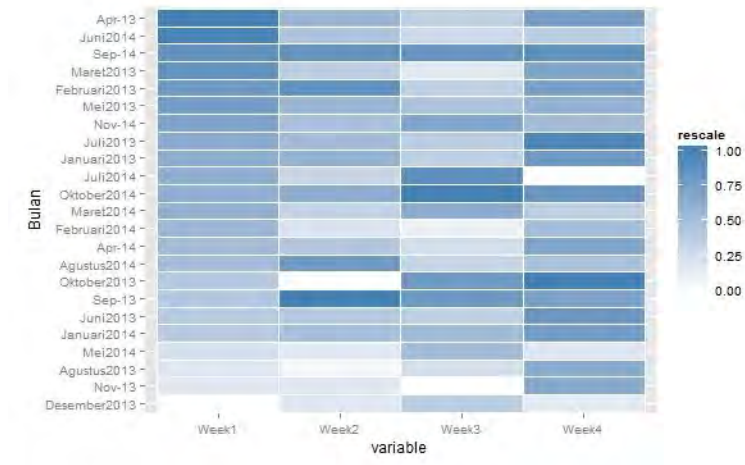
Untuk PT. XYZ memerlukan laporan bagaimana transaksi per minggu, dan pada minggu ke berapa dan bulan apa pelanggan paling banyak melakukan transaksi dalam jumlah banyak.

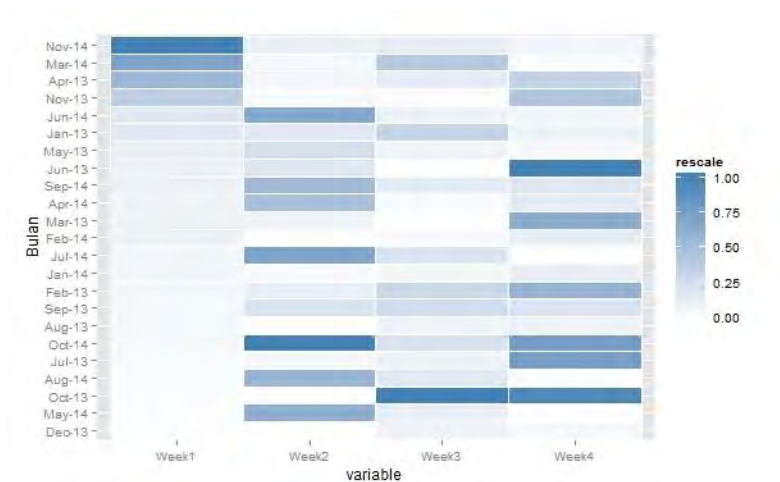
Laporan tersebut akan digunakan oleh PT. XYZ untuk membandingkannya dengan catatan medis atau kejadian yang berkaitan tentang kesehatan untuk mempersiapkan strategi pemasaran jika kondisi seperti itu terjadi lagi.

Analisis *behavior* ini akan dibuat dengan bantuan *tools* aplikasi R. Cara untuk membuatnya telah dibahas pada bab 2.2.8.3. Karena skala yang ditunjukkan adalah 0 sampai 1, skala 0 di sini menunjukkan kalau nilai itu minimal dari klaster ini dan 1 maksimal dari klaster ini. Tidak berarti skala 0 menunjukkan angka 0 juga pada nilai asli.

Hasil analisis *behavior* dari pelanggan yang berada pada klaster 1 yang bisa dilihat pada tabel 6.1 dan 6.2.

**Tabel 6.8 Transaksi klaster 1 per minggu**



**Tabel 6.9 Skala transaksi klaster 1 per minggu**

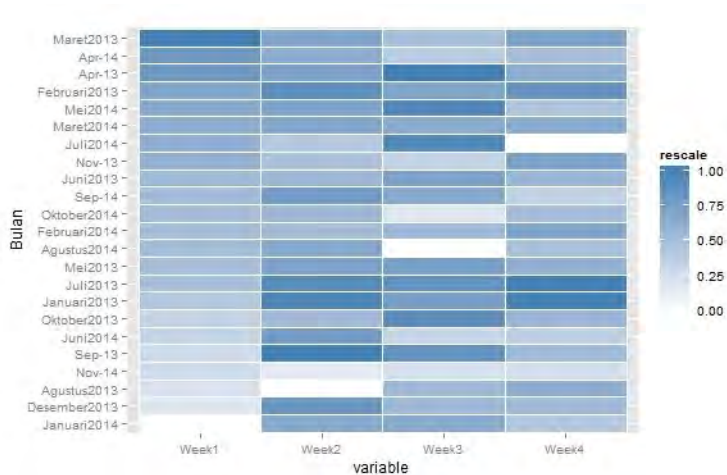
Beberapa informasi yang bisa diambil dari hasil analisis klaster 4 pada tabel 6.1 dan 6.2 adalah:

- Pelanggan pada klaster 1 ini paling sering melakukan transaksi pada 4 bulan awal 2013.
- Pelanggan pada klaster 1 ini sering melakukan transaksi pada minggu ketiga setiap bulannya.
- Pelanggan pada klaster ini paling banyak melakukan transaksi pada minggu kedua bulan September 2013 dan paling sedikit pada bulan Juli 2014 minggu keempat.
- Pelanggan pada klaster ini paling banyak melakukan transaksi dalam jumlah besar pada bulan Oktober 2014 di minggu kedua dan paling sedikit pada bulan Mei 2014 minggu keempat.

**Tabel 6.10 Nilai Ekstrim pada Klaster 1**

Analisis	Nilai pada Minggu Ekstrim	Minggu			
		1	2	3	4
Frequency	5653				
Monetary	5613630025				

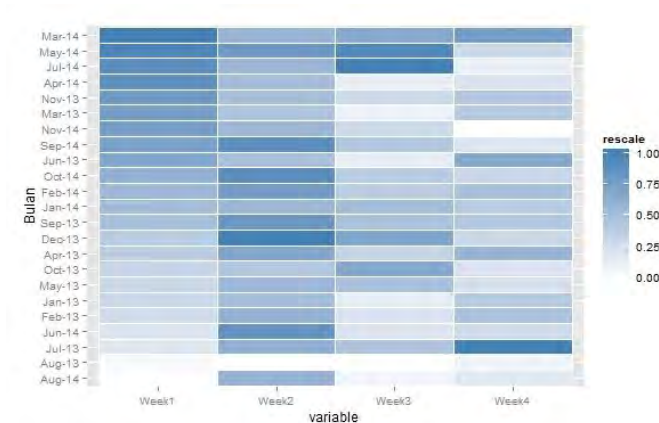
Hasil analisis *behavior* dari pelanggan yang berada pada klaster 2 yang bisa dilihat pada tabel 6.3 dan 6.4.

**Tabel 6.11 Transaksi klaster 2 per minggu****Tabel 6.12 Nilai Ekstrim pada Klaster 2**

Analisis	Nilai pada Minggu Ekstrim	Minggu			
		1	2	3	4
Frequency	21182				
Monetary	20690919998				



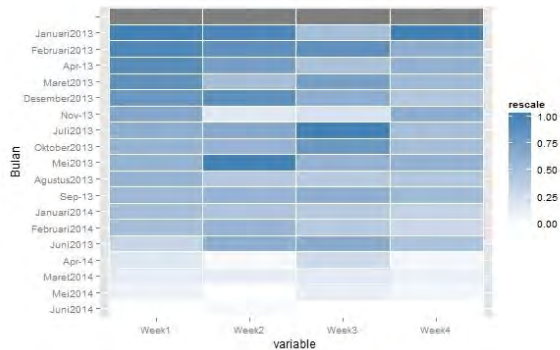
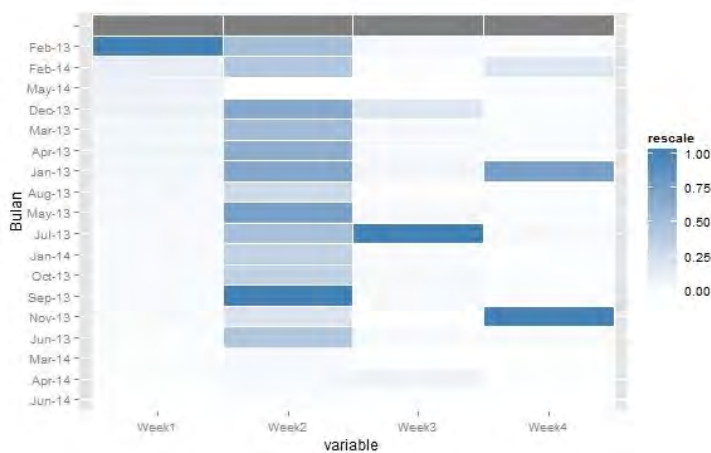
Tabel 6.13 Skala transaksi klaster 2 per minggu



Beberapa informasi yang bisa diambil dari hasil analisis klaster 2 pada tabel 6.3 dan 6.4 adalah:

- Pelanggan pada klaster 2 ini sering melakukan transaksi pada minggu ketiga setiap bulannya.
- Pelanggan pada klaster ini paling banyak melakukan transaksi pada minggu keempat bulan Juli 2013 dan paling sedikit pada bulan Juli 2014 minggu keempat.
- Pelanggan pada klaster ini paling banyak melakukan transaksi dalam jumlah besar pada bulan Juli 2013 di minggu keempat dan paling sedikit pada bulan Agustus 2013 minggu kedua.

Hasil analisis *behavior* dari pelanggan yang berada pada klaster 3 yang bisa dilihat pada tabel 6.5 dan 6.6.

**Tabel 6.14 Transaksi klaster 3 per minggu****Tabel 6.15 Skala transaksi klaster 3 per minggu**

Beberapa informasi yang bisa diambil dari hasil analisis klaster 3 pada tabel 6.5 dan 6.6 adalah:

- Pelanggan pada klaster 3 berhenti melakukan transaksi pada bulan Juli 2014 minggu ketiga.
- Pelanggan pada klaster ini juga mulai menurun melakukan transaksinya saat bulan Februari 2014.

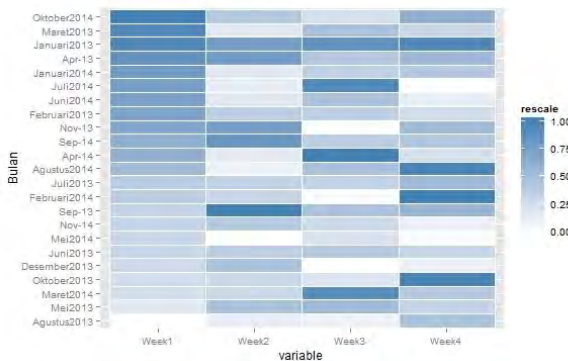
- Pelanggan pada klaster 3 ini sering melakukan transaksi pada minggu ketiga setiap bulannya.
- Pelanggan pada klaster ini paling banyak melakukan transaksi pada minggu ketiga bulan Juli 2013 dan paling sedikit pada bulan Juni 2014 minggu ketiga.
- Pelanggan pada klaster ini paling banyak melakukan transaksi dalam jumlah besar pada bulan November 2013 di minggu keempat dan paling sedikit pada bulan Maret 2014 minggu keempat.

**Tabel 6.16 Nilai Ekstrim pada Klaster 3**

Analisis	Nilai pada Minggu Ekstrim	Minggu			
		1	2	3	4
Frequency	303				
Monetary	267803445				

Untuk pelanggan yang berada pada klaster 4 maka hasil analisisnya bisa dilihat pada tabel 6.7 dan 6.8.

**Tabel 6.17 Transaksi klaster 4 per minggu**



**Tabel 6.18 Skala transaksi kluster 4 per minggu**

Beberapa informasi yang bisa diambil dari hasil analisis kluster 4 pada tabel 6.7 dan 6.8 adalah:

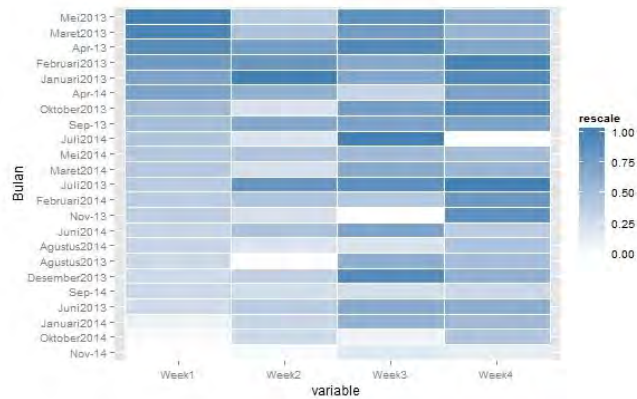
- Pelanggan pada kluster 4 ini sering melakukan transaksi pada minggu pertama setiap bulannya.
- Pelanggan pada kluster ini paling banyak melakukan transaksi pada minggu pertama bulan Oktober 2014 dan paling sedikit pada bulan Juni 2014 minggu keempat.
- Pelanggan pada kluster ini paling banyak melakukan transaksi dalam jumlah besar pada bulan Juni 2014 di minggu pertama dan paling sedikit pada bulan Agustus 2013 minggu pertama.

**Tabel 6.19 Nilai Ekstrim pada Kluster 4**

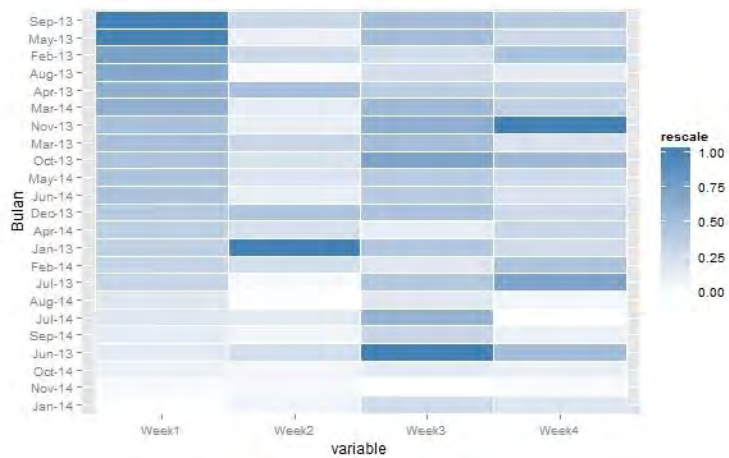
Analisis	Nilai pada Minggu Ekstrim	Minggu			
		1	2	3	4
Frequency	1275				
Monetary	1582653587				

Hasil analisis *behavior* dari *cluster* 5 terdapat pada tabel 6.9 dan 6.10.

Tabel 6.20 Transaksi klaster 5 per minggu



Tabel 6.21 Skala transaksi klaster 5 per minggu



Berdasarkan tabel 6.9 dan 6.10 maka beberapa kesimpulan yang bisa diambil adalah:

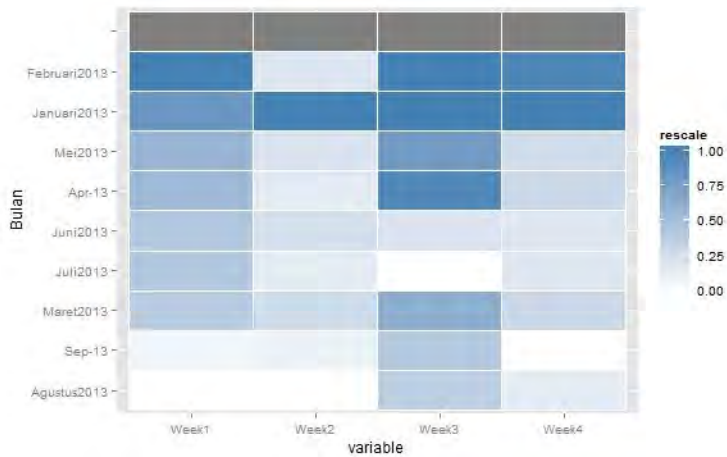
- Pelanggan pada klaster 5 paling banyak melakukan transaksi pada semester 5 (Januari-Juli)
- Pelanggan pada klaster 5 transaksi sangat jarang melakukan transaksi pada semester kedua terakhir (Juli-Desember)
- Pelanggan pada klaster 5 ini sering melakukan transaksi pada minggu ketiga setiap bulannya.
- Pelanggan pada klaster ini paling banyak melakukan transaksi pada minggu kedua bulan Januari 2013 dan paling sedikit pada bulan Jui 2014 minggu keempat.
- Pelanggan pada klaster ini paling banyak melakukan transaksi dalam jumlah besar pada bulan Januari 2013 di minggu kedua dan paling sedikit pada bulan Agustus 2014 minggu kedua.

**Tabel 6.22 Nilai Ekstrim pada Klaster 5**

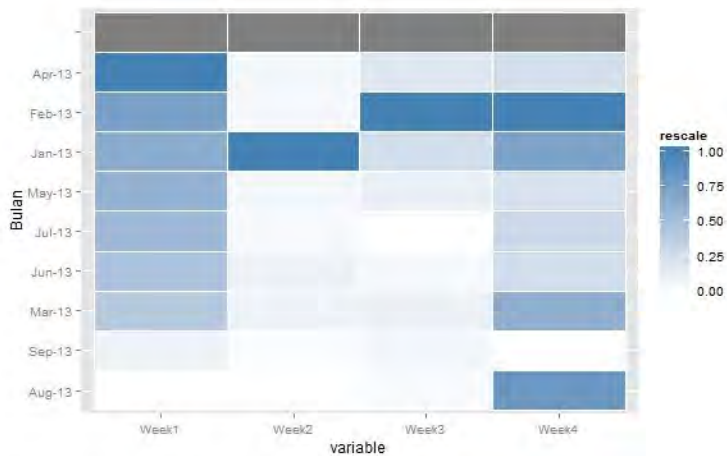
Analisis	Nilai pada Minggu Ekstrim	Minggu			
		1	2	3	4
Frequency	8870				
Monetary	5881818865				

Selanjutnya adalah hasil analisis *behavior* dari klaster 6 yang bisa dilihat pada tabel 6.11 dan 6.12.

Tabel 6.23 Transaksi klaster 6 per minggu



Tabel 6.24 Skala transaksi klaster 6 per minggu



Informasi yang bisa diambil dari tabel 6.11 dan 6.12 adalah:

- Pelanggan pada klaster 6 ini hanya melakukan transaksi terbanyak pada bulan Januari 2013 dan kemudian menurun.
- Pelanggan klaster 6 hanya melakukan transaksi hingga bulan September tahun 2013 minggu keempat.
- Pelanggan pada klaster 6 ini sering melakukan transaksi pada minggu kedua setiap bulannya.
- Pelanggan pada klaster ini paling banyak melakukan transaksi pada minggu pertama bulan Februari 2013 dan paling sedikit pada bulan Juli 2013 minggu ketiga dan Agustus 2013 pada minggu kedua keduanya berjumlah 0 transaksi.
- Pelanggan pada klaster ini paling banyak melakukan transaksi dalam jumlah besar pada bulan Januari 2013 di minggu kedua dan paling sedikit pada bulan Juli 2013 minggu ketiga dan Agustus 2013 minggu kedua.

**Tabel 6.25 Nilai Ekstrim pada Klaster 6**

Analisis	Nilai pada Minggu Ekstrim	Minggu			
		1	2	3	4
Frequency	140				
Monetary	46830700				

## 6.2. Menemukan Karakteristik Pelanggan

Setelah melakukan analisis geografis, demografis *behavior* dan atribut RFM maka selanjutnya adalah langkah untuk menentukan karakteristik dari setiap klaster pelanggan dengan menggabungkan keempat hasil analisis dengan studi literature yang telah dilakukan. Karakteristik pelanggan dapat dilihat seperti tabel 6.16 di bawah ini.



Tabel 6.26 Tabel karakteristik pelanggan

Cluster	Karakteristik	Label Pelanggan
1	<ul style="list-style-type: none"> <li>- Pelanggan yang <b>lumayan sering</b> melakukan transaksi hingga bulan terakhir dilakukan analisis. Namun <b>tidak</b> membeli produk/barang dalam <b>skala banyak</b>.</li> <li>- Pelanggan berada agak tersebar di luar kota Semarang.</li> <li>- Pelanggan pada klaster 1 ini sering melakukan transaksi pada minggu ketiga setiap bulannya.</li> <li>- Pelanggan pada klaster ini berjumlah 107 pelanggan atau sekitar 22,5% dari total pelanggan.</li> </ul>	<i>Shoppers</i>
2	<ul style="list-style-type: none"> <li>- Pelanggan ada klaster 2 ini adalah pelanggan yang <b>paling loyal</b> karena <b>sangat sering</b> melakukan transaksi dan juga pembelian yang dilakukan rata-rata <b>dalam skala yang besar</b>.</li> <li>- Pelanggan mayoritas adalah <b>apotik</b> dan <b>rumah sakit</b>.</li> <li>- Pelanggan klaster ini berada tersebar di luar kota</li> <li>- Pelanggan pada klaster 2 ini sering melakukan transaksi pada minggu ketiga setiap bulannya.</li> <li>- Sebanyak 8% pelanggan berada di</li> </ul>	<i>Best</i>

Cluster	Karakteristik	Label Pelanggan
	klaster 2.	
3	<ul style="list-style-type: none"> <li>- Merupakan pelanggan yang <b>sudah lama</b> tidak melakukan transaksi, <b>jarang</b> melakukan transaksi dan belanjanya juga <b>sedikit</b>.</li> <li>- Mayoritas pelanggan <b>berkumpul</b> di kota Semarang.</li> <li>- Pelanggan ini hanya melakukan transaksi sampai bulan Juli 2014 dan sudah mulai menurun saat bulan Februari 2014.</li> <li>- Terdapat 12% pelanggan terdapat pada klaster ini.</li> </ul>	<i>Uncertain</i>
4	<ul style="list-style-type: none"> <li>- Pelanggan pada klaster ini <b>cukup sering</b> melakukan transaksi sampai akhir-akhir ini namun <b>tidak dalam skala besar</b>.</li> <li>- Lokasi pelanggan banyak yang berada pada kota Semarang.</li> <li>- Pelanggan paling banyak melakukan transaksi pada minggu pertama.</li> <li>- Presentase pelanggan yang berada pada klaster ini adalah 34,5%.</li> </ul>	<i>First Time</i>
5	<ul style="list-style-type: none"> <li>- Pelanggan pada klaster ini pelanggan yang cukup baik karena <b>sering</b> melakukan transaksi sampai akhir-akhir ini dan transaksinya juga dalam</li> </ul>	<i>Valueable</i>

Cluster	Karakteristik	Label Pelanggan
	<p>jumlah yang <b>cukup besar</b>.</p> <ul style="list-style-type: none"> <li>- Pelanggan berada pada daerah kota Semarang.</li> <li>- Pelanggan sering melakukan transaksi pada minggu ketiga setiap bulannya.</li> <li>- Pelanggan pada klaster ini ada 12% dari total keseluruhan pelanggan</li> </ul>	
6	<ul style="list-style-type: none"> <li>- Pelanggan pada klaster ini adalah pelanggan yang paling buruk karena dari ketigaatributnya tidak ada yang baik. <b>Sudah lama dan sangat jarang melakukan transaksi</b>, dan juga saat melakukan transaksi skalanya <b>sangat kecil</b>.</li> <li>- Mayoritas pelanggan juga berada sangat jauh dari kota Semarang.</li> <li>- Pelanggan hanya melakukan transaksi sampai minggu keempat bulan September 2013</li> <li>- Terdapat 10% pelanggan terdapat pada klaster ini.</li> </ul>	<i>Churn</i>

Berdasarkan hasil yang terlihat pada tabel 6.8 maka didapatkan karakteristik dan tipe pelanggan dari keenam segmen pelanggan yang didapatkan dari proses *clustering*.

## **BAB VII**

### **KESIMPULAN DAN SARAN**

Bab ini berisikan kesimpulan dari hasil penelitian dan juga saran perbaikan untuk penelitian kedepannya beserta masalah yang dihadapi selama mengerjakan penelitian tugas akhir ini.

#### **7.1. Kesimpulan**

Beberapa kesimpulan yang bisa diambil dari penelitian tugas akhir ini adalah:

1. Berdasarkan hasil *clustering* dengan metode Fuzzy Clustering Method jumlah segmen pelanggan yang terbentuk dari pelanggan-pelanggan PT. XYZ adalah 6 klaster/segmen pelanggan dari keseluruhan jumlah pelanggan sebanyak 475 pelanggan. Dengan nilai SSE 4,988 dan nilai DBI 0.57
2. Sejumlah 475 pelanggan tersebut terbagi ke dalam enam kelompok berdasarkan nilai variabel RFM oleh algoritma Fuzzy C-Means, masing-masing kelompok tersebut memiliki karakteristik sebagai berikut:
  - a. Klaster 1 merupakan klaster dengan label *shoppers*. Pelanggan pada klaster ini melakukan transaksi terakhir pada jangka waktu dekat, dengan frekuensi transaksi di atas rata-rata, dan besar transaksinya di bawah rata-rata. Pelanggan pada klaster ini berada agak tersebar di luar kota Semarang. Pelanggan pada klaster ini juga sering melakukan transaksi pada minggu ketiga setiap bulannya.
  - b. Klaster 2 merupakan klaster dengan label *best*. Pelanggan pada klaster ini melakukan transaksi terakhir pada jangka waktu dekat, dengan frekuensi transaksi, dan besar transaksinya di atas rata-rata. Pelanggan pada klaster ini berada tersebar di luar kota Semarang. Pelanggan pada klaster ini sering melakukan transaksi pada minggu ketiga.
  - c. Klaster 3 merupakan klaster dengan label *uncertain*. Pelanggan pada klaster ini melakukan transaksi terakhir pada jangka waktu lama, dengan frekuensi transaksi, dan besar transaksinya di bawah rata-rata. Pelanggan pada

- klaster ini berada terkumpul di kota Semarang. Pelanggan ini hanya melakukan transaksi sampai bulan Juli 2014.
- d. Klaster 4 merupakan klaster dengan label *first time*. Pelanggan pada klaster ini melakukan transaksi terakhir pada jangka waktu dekat, dengan frekuensi transaksi, dan besar transaksinya di bawah rata-rata. Pelanggan pada klaster ini berada terkumpul di dalam kota Semarang. Pelanggan pada klaster ini juga sering melakukan transaksi pada minggu pertama setiap bulannya.
  - e. Klaster 5 merupakan klaster dengan label *valuable*. Pelanggan pada klaster ini melakukan transaksi terakhir pada jangka waktu dekat, dengan frekuensi transaksi, dan besar transaksinya di atas rata-rata. Pelanggan pada klaster ini berada terkumpul di kota Semarang. Pelanggan pada klaster ini juga sering melakukan transaksi pada minggu ketiga setiap bulannya.
  - f. Klaster 6 merupakan klaster dengan label *Churn*. Pelanggan pada klaster ini melakukan transaksi terakhir pada jangka waktu lama, dengan frekuensi transaksi, dan besar transaksinya di bawah rata-rata. Pelanggan pada klaster ini berada tersebar di luar kota Semarang. Pelanggan pada klaster ini sudah tidak melakukan transaksi pada minggu keempat bulan September 2013.
3. Terdapat dua klaster pelanggan yang mempunyai nilai RFM sangat baik dan sangat buruk. Klaster pelanggan yang mempunyai nilai RFM paling baik berada pada klaster 2 yang mempunyai anggota 38 pelanggan dengan label *best*. Sedangkan untuk klaster yang mempunyai nilai RFM terburuk adalah klaster 6 yang mempunyai anggota 49 pelanggan dengan label *Churn*.
  4. Berdasarkan analisis tipe pelanggan, pelanggan yang perlu diperhatikan lebih berturut-turut adalah sebagai berikut:
    - a. Rumah Sakit. Tipe pelanggan rumah sakit terdapat pada nilai RFM terbaik sebanyak 39%, kemudian pada nilai RFM kedua terbaik sejumlah 23%, untuk nilai RFM ketiga

terbaik terdapat 20% rumah sakit, dan untuk nilai RFM terburuk ada sejumlah 14% rumah sakit.

- b. Apotik. Tipe pelanggan apotik terdapat pada nilai RFM terbaik sebanyak 19%, kemudian pada nilai RFM kedua terbaik sejumlah 23%, untuk nilai RFM ketiga terbaik terdapat 35% apotik, dan untuk nilai RFM terburuk ada sejumlah 23% apotik.
- c. Klinik. Tipe pelanggan klinik terdapat pada nilai RFM terbaik sebanyak 0%, kemudian pada nilai RFM kedua terbaik sejumlah 0%, untuk nilai RFM ketiga terbaik terdapat 75% klinik, dan untuk nilai RFM terburuk ada sejumlah 25% klinik.
- d. Perusahaan/Yayasan. Tipe pelanggan perusahaan/yayasan terdapat pada nilai RFM terbaik sebanyak 0%, kemudian pada nilai RFM kedua terbaik sejumlah 0%, untuk nilai RFM ketiga terbaik terdapat 45% perusahaan/yayasan, dan untuk nilai RFM terburuk ada sejumlah 55% perusahaan/yayasan.

## 7.2. Saran

Saran yang dapat dipertimbangkan untuk pengembangan tugas akhir ini yaitu:

1. Perlu ditambahkan atribut selain *recency*, *frequency*, dan *monetary* yang digunakan untuk menentukan segmentasi pelanggan dengan atribut *Length* untuk menentukan lama keanggotaan suatu pelanggan, *Cost* untuk memasukkan variabel biaya atau *Location* untuk menambahkan lokasi atau jarak dari perusahaan.
2. Memaksimalkan visualisasi untuk membantu para pengambil keputusan menganalisis segmen pelanggan. Visualisasi yang dimaksud diharapkan dapat dinamis, sehingga apabila dimasukkan data baru secara otomatis menampilkan visualisasi.
3. Analisis yang dilakukan seharusnya lebih mendalam lagi dengan menambahkan analisis demografinya dengan apakah pelanggan termasuk perusahaan besar atau kecil.

## DAFTAR PUSTAKA

- [1] Tim Pengelola Website Kemenperin. (2011) Kementerian Perindustrian Republik Indonesia. [Online]. <http://www.kemenperin.go.id/artikel/2790/Industri-Farmasi-Minta-Tax-Holiday>
- [2] Tim Dunia Informasi Kesehatan. (2014) Dunia Informasi Kesehatan. [Online]. <http://www.duniainformasikesehatan.com/2014/03/manfaat-kesehatan-bagi-hidup-kita.html>
- [3] JR Miglautchs, "Thoughts On RFM Scoring," *Journal of Database Marketing* 8, vol. 8, pp. 67-72.
- [4] S Chaitanya Sripada and S Rao, "COMPARISON OF PURITY AND ENTROPY OF K-MEANS CLUSTERING AND FUZZY C MEANS CLUSTERING," *Indian Journal of Computer Science and Engineering*, pp. 343-346, 2011.
- [5] Adrian Payne and Pennie Frow, "A Strategic Framework for Customer," *Journal of Marketing*, pp. 167-177, 2005.
- [6] T. L. Oldano, "Relationship segmentation. Enhancing the service provider/client connection," in *Add Value to Your Service. The Key to Success, Proceedings Series*. Chicago: AMA, 1987, pp. 143-146.
- [7] Michael J.A Berry, *Data Mining Techniques for CRM*. Canada: Wiley Publishing, 2004.
- [8] Andrei Scridon Mircea, "UNDERSTANDING CUSTOMERS - PROFILING AND SEGMENTATION," *Management & Marketing*, no. 1, pp. 175-184, 2008.
- [9] Bart J. Bronnenberg and Paulo Albuquerque, "Geography and Marketing Strategy in Consumer Packaged Goods," *Geography and Marketing Strategy in Consumer*, vol. Third, Desember 2002.
- [10] M. Fujita, P. Krugman, and A. J. Venables, *The Spatial*

- Economy*. Cambridge, Massachusetts: The MIT Press.
- [11] Canada Business Network. (2009) INFO ENTREPRENEURS. [Online].  
<http://www.infoentrepreneurs.org/en/guides/know-your-customers--needs/>
- [12] Jim Blythe and Phil Megicks, *Marketing Planning*.: Prentice Hall, 2013.
- [13] J. R. Bult and Wansbeek T., "Optimal selection for direct mail," *Marketing Science*, vol. 14, pp. 378-394, 1995.
- [14] R. C. Blattberg, B-D. Kim, and S. A. Neslin, "Chapter 12," in *Database Marketing: Analyzing and Managing*. New York: Springer, 2008, pp. 323-337.
- [15] Derya Birant, "Data Mining Using RFM Analysis," in *Knowledge-Oriented Applications in Data Mining*.: InTech, 2011.
- [16] Efraim Turban, Ramesh Sharda, and Dursun Delen, *Decision Support and Business Intelligence Systems 9th Edition*. New Jersey: Pearson, 2011.
- [17] Florin Goronescu, *Data Mining - Concept, Model, and Techniques*. Berlin: Springer - Verlag Verlin Heidelberg, 2011.
- [18] Efraim Turban, Ramesh Sharda, and Dursun Delen, *Decision Support and Business Intelligence Systems 9th Edition*. New Jersey: Pearson, 2011.
- [19] Eko Prasteyo, "Konsep Cluster," in *DATA MINING. Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: ANDI, 2014, pp. 13-15.
- [20] Eko Prasetyo, *Data Mining - Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi Offset, 2013.
- [21] J. C. Bezdek, R. Erlich, and W. Full, "FCM : The Fuzzy C-Means Clustering Algorithm," *Computer & Geoscience*, pp. 1-13, 1981.



- [22] (2015) Testing and Validation (Data Mining). [Online].  
<https://msdn.microsoft.com/en-us/library/ms174493.aspx>
- [23] Joshua Knowles and Douglas Kell Julia Handl. (2005) Cluster validation. [Online].  
<http://personalpages.manchester.ac.uk/mbs/julia.handl/validation.html>
- [24] L. Rokach and O. Maimon, *CLUSTERING METHODS. DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK.*, 2009, pp. 330-343.
- [25] Juan Carlos Rojas Thomas, "New Version of Davies-Bouldin Index for Clustering," p. 5.
- [26] Nursinta A.Wahanani, Aziz Kustiyo Yandra Arkeman, "Clustering K-Means Optimization with Multi-Objective Genetic Algorithm," *International Journal of Electrical & Computer Sciences*, p. 6.
- [27] W. N. Venables, D. M. Smith, and the R Core Team, *An Introduction to R*. Adelaide, Australia, 2014.
- [28] M Hardin, D Hom, R Perez, and L Williams, "Which chart or graph is right for you?," *Whitepaper Tableau*, pp. 10-15, 2013.
- [29] K. C Hsien, "Bivariate and multivariate data and distribution.," 2006.

## LAMPIRAN A

### DATA TRANSAKSI PELANGGAN

Tabel A 1 Cuplikan Data Transaksi Pelanggan

BLN	THN	KODE LANGG	TGL	KODE PROD	UNIT	hna pusat	tot by pusat
01	2013	95854	3	TKNXK	1	234000	234000
01	2013	95854	3	TZEGA	3	100000	300000
01	2013	95854	4	AKMEB	10	26500	265000
01	2013	95854	4	AKMED	20	25000	500000
01	2013	95799	5	VCRVD	6	290000	1740000
01	2013	95854	7	KCLMB	1	309000	309000
01	2013	95854	7	TKNXK	2	234000	468000
01	2013	95854	7	TKTFA	2	72000	144000
01	2013	95854	7	TKTFB	2	125000	250000
01	2013	95854	7	TZEGA	2	100000	200000
01	2013	95854	7	VCRVA	10	250000	2500000
01	2013	95854	9	VTIVC	6	62000	372000
01	2013	86787	9	TKNXK	1	234000	234000

BLN	THN	KODE LANGG	TGL	KODE PROD	UNIT	hna pusat	tot by pusat
01	2013	86787	9	TZEGA	1	100000	100000
01	2013	95854	10	CKTFA	3	100000	300000
01	2013	95854	10	TCOXK	3	297000	891000
01	2013	95799	10	VTIVC	4	62000	248000
01	2013	95854	11	TCEFE	5	264000	1320000
01	2013	95854	12	TZEGA	3	100000	300000
01	2013	95799	12	AKNXB	20	100800	2016000
01	2013	95854	14	IHXLA	20	73500	1470000
01	2013	95854	14	TCRVK	10	315000	3150000
01	2013	95799	14	AKNXB	100	100800	10080000
01	2013	95799	14	AKNXB	100	100800	10080000
01	2013	95854	16	TKNXK	2	234000	468000
01	2013	95799	16	AKNXB	140	100800	14112000
01	2013	95799	16	CKTFA	10	100000	1000000
01	2013	95799	16	KCEFB	11	510000	5610000
01	2013	95799	16	KGSDA	354	142500	50445000
01	2013	95799	16	TCEFE	20	264000	5280000

<b>BLN</b>	<b>THN</b>	<b>KODE LANGG</b>	<b>TGL</b>	<b>KODE PROD</b>	<b>UNIT</b>	<b>hna pusat</b>	<b>tot by pusat</b>
01	2013	95799	16	TKNXK	9	234000	2106000
01	2013	95799	16	TZEGA	45	100000	4500000
01	2013	95799	16	VCOXE	68	79500	5406000
01	2013	95799	16	VCRVD	20	290000	5800000
01	2013	95854	17	KGSDA	10	142500	1425000
01	2013	95854	17	KVTIA	5	70000	350000
01	2013	95799	18	TCRVE	5	210000	1050000
01	2013	95854	21	KVTIA	5	70000	350000
01	2013	95854	21	TKTFA	2	72000	144000
01	2013	86787	21	KGSDA	4	142500	570000
01	2013	86787	19	VTIVC	2	62000	124000
01	2013	95854	23	AKMED	10	25000	250000
01	2013	95854	23	KCEFA	1	245000	245000
01	2013	95854	23	KCEFB	10	510000	5100000
01	2013	86787	23	TKNXK	2	234000	468000
01	2013	95854	25	AKNXB	10	100800	1008000
01	2013	95854	25	VCRVA	10	250000	2500000

BLN	THN	KODE LANGG	TGL	KODE PROD	UNIT	hna pusat	tot by pusat
01	2013	95854	26	KGSDA	10	142500	1425000
01	2013	95854	28	CKTFA	2	100000	200000
01	2013	95854	28	KGSDA	20	142500	2850000
01	2013	95854	28	KUNCA	2	456000	912000
01	2013	95854	28	TCEFE	10	264000	2640000
01	2013	95854	28	TKNXK	1	234000	234000
01	2013	95854	28	TZEGA	1	100000	100000
01	2013	95799	28	VTIVC	10	62000	620000
01	2013	95854	31	AKNXB	20	100800	2016000
01	2013	95854	31	KVTIA	10	70000	700000
01	2013	95854	31	TCRVK	10	315000	3150000
01	2013	95854	31	TZEGA	3	100000	300000
01	2013	95854	31	VCRVA	10	250000	2500000
01	2013	95854	31	VTIVC	6	62000	372000
01	2013	95854	3	ICFZE	5	82000	410000
01	2013	95854	3	IVPMA	5	259000	1295000
01	2013	95854	3	VBROK	20	164500	3290000

BLN	THN	KODE LANGG	TGL	KODE PROD	UNIT	hna pusat	tot by pusat
01	2013	95854	4	VTHDB	10	164000	1640000
01	2013	95854	7	IMIKB	30	155000	4650000
01	2013	95854	7	ITRSB	5	185000	925000
01	2013	95854	7	VBROK	30	164500	4935000
01	2013	95854	7	VCNEA	100	107000	10700000
01	2013	95854	9	VTHDB	20	164000	3280000
01	2013	95854	10	VBROK	20	164500	3290000
01	2013	95854	12	IMIKB	50	155000	7750000
01	2013	95854	12	TTRSA	2	97000	194000
01	2013	95854	12	VBROK	30	164500	4935000
01	2013	95854	12	VKFXP	10	115000	1150000
01	2013	95854	14	ITRSB	5	185000	925000
01	2013	95799	16	ICFXG	40	196000	7840000
01	2013	95799	16	IMRFC	60	420000	25200000
01	2013	95799	16	ITRSB	16	185000	2960000
01	2013	95854	21	IMIKB	50	155000	7750000
01	2013	95799	22	VCNEA	130	107000	13910000

BLN	THN	KODE LANGG	TGL	KODE PROD	UNIT	hna pusat	tot by pusat
01	2013	95854	23	VKMYP	20	23000	460000
01	2013	95854	28	VBROK	30	164500	4935000
01	2013	95854	3	LLKMA	10	27000	270000
01	2013	86787	4	LFKDA	2	30000	60000
01	2013	86787	5	GDLCA	1	96000	96000
01	2013	95854	7	DCEFA	20	80000	1600000
01	2013	95854	7	KFCLB	1	580000	580000
01	2013	95854	7	TPFTA	3	115000	345000
01	2013	95854	7	TPPGA	1	46000	46000
01	2013	95854	8	AIDXB	5	75000	375000
01	2013	95854	8	TFROE	3	145000	435000
01	2013	95854	8	TVMAA	10	21000	210000
01	2013	86787	9	TPFLA	1	125000	125000
01	2013	95854	10	LENSA	6	25000	150000
01	2013	95854	10	LLKMA	10	27000	270000
01	2013	95854	10	TPFTA	3	115000	345000
01	2013	95854	10	TPPGA	2	46000	92000

## LAMPIRAN B

### NILAI RFM PELANGGAN

**Tabel A 2 Cuplikan Nilai RFM Pelanggan**

Kode Langganan	Freq	Mon	Rec	FreqNorm	MonNorm	RecNorm
902347	85	18289500	82	0.1484	0.0009	0.0043
9023594	139	26250840	87	0.2438	0.0013	0.0116
96033	129	43009170	80	0.2261	0.0022	0.0014
903953	64	13020500	95	0.1113	0.0007	0.0232
515049	49	7225000	82	0.0848	0.0004	0.0043
86848	92	16985000	91	0.1608	0.0009	0.0174
730847	58	10173000	91	0.1007	0.0005	0.0174
86626	78	49458000	88	0.1360	0.0025	0.0130
96021	69	12710250	81	0.1201	0.0006	0.0029
52835	51	18288700	80	0.0883	0.0009	0.0014
86696	111	57576250	81	0.1943	0.0029	0.0029
86611	69	22454000	84	0.1201	0.0011	0.0072
86472	85	14397400	80	0.1484	0.0007	0.0014
86878	64	7651000	93	0.1113	0.0004	0.0203
929642	118	352125700	79	0.2067	0.0177	0.0000
593708	48	17335400	79	0.0830	0.0009	0.0000
579291	58	10412500	97	0.1007	0.0005	0.0260
86621	65	10512050	124	0.1131	0.0005	0.0651
9023593	116	34598280	103	0.2032	0.0017	0.0347
96008	136	25955850	94	0.2385	0.0013	0.0217
557173	76	25108000	79	0.1325	0.0013	0.0000
920550	77	13793700	83	0.1343	0.0007	0.0058
86789	61	15400980	91	0.1060	0.0008	0.0174



Kode Langganan	Freq	Mon	Rec	FreqNorm	MonNorm	RecNorm
629472	85	14525500	80	0.1484	0.0007	0.0014
95955	126	30904205	87	0.2208	0.0015	0.0116
86836	131	347765900	81	0.2297	0.0174	0.0029
86690	93	13128000	81	0.1625	0.0007	0.0029
9023592	72	14597500	80	0.1254	0.0007	0.0014
96031	122	75097000	83	0.2138	0.0038	0.0058
599074	84	13441000	91	0.1466	0.0007	0.0174
118854	124	15329350	79	0.2173	0.0008	0.0000
86732	127	146480385	79	0.2226	0.0073	0.0000
902348	77	18755400	88	0.1343	0.0009	0.0130
144515	134	343784875	79	0.2350	0.0172	0.0000
86553	103	34383000	93	0.1802	0.0017	0.0203
96029	83	16544000	84	0.1449	0.0008	0.0072
902349	137	31220500	133	0.2403	0.0016	0.0781
86613	147	26903500	266	0.2580	0.0013	0.2706
86565	121	25119400	133	0.2120	0.0013	0.0781
86911	112	21236500	242	0.1961	0.0011	0.2359
831771	151	33027290	84	0.2650	0.0017	0.0072
9023591	123	20081360	79	0.2155	0.0010	0.0000
96012	78	16343700	112	0.1360	0.0008	0.0478
720914	128	31604000	79	0.2244	0.0016	0.0000
119056	110	20174850	83	0.1926	0.0010	0.0058
902346	75	8395500	90	0.1307	0.0004	0.0159
531569	75	18940500	81	0.1307	0.0009	0.0029
927650	54	10537750	97	0.0936	0.0005	0.0260
86709	115	18534130	80	0.2014	0.0009	0.0014
614869	86	22149700	81	0.1502	0.0011	0.0029
449826	122	20342500	83	0.2138	0.0010	0.0058

Kode Langganan	Freq	Mon	Rec	FreqNorm	MonNorm	RecNorm
639096	89	24862000	125	0.1555	0.0012	0.0666
622352	85	36302500	95	0.1484	0.0018	0.0232
86575	88	41174500	80	0.1537	0.0021	0.0014
577589	95	29386265	80	0.1661	0.0015	0.0014
733609	139	20030700	82	0.2438	0.0010	0.0043
86902	94	22240600	87	0.1643	0.0011	0.0116
504739	71	13292400	83	0.1237	0.0007	0.0058
86895	81	15304000	87	0.1413	0.0008	0.0116
571945	55	9278800	81	0.0954	0.0005	0.0029
96033	89	17786500	82	0.1555	0.0009	0.0043
9023590	77	9427000	86	0.1343	0.0005	0.0101
9023589	145	35244685	79	0.2544	0.0018	0.0000
86563	83	20015735	91	0.1449	0.0010	0.0174
95783	97	19799900	81	0.1696	0.0010	0.0029
86797	56	5483500	108	0.0972	0.0003	0.0420
86779	54	10753250	83	0.0936	0.0005	0.0058
744905	70	21703140	87	0.1219	0.0011	0.0116
575033	63	15606750	81	0.1095	0.0008	0.0029
86705	128	32375750	83	0.2244	0.0016	0.0058
86699	89	29766000	87	0.1555	0.0015	0.0116
536135	124	33696430	87	0.2173	0.0017	0.0116
620833	144	50189250	81	0.2527	0.0025	0.0029
86830	111	35374350	79	0.1943	0.0018	0.0000
86865	111	16030000	91	0.1943	0.0008	0.0174
86491	69	12386500	94	0.1201	0.0006	0.0217
86622	67	9729500	101	0.1166	0.0005	0.0318
86719	66	14761850	86	0.1148	0.0007	0.0101
474893	130	113911000	80	0.2279	0.0057	0.0014

<b>Kode Langganan</b>	<b>Freq</b>	<b>Mon</b>	<b>Rec</b>	<b>FreqNorm</b>	<b>MonNorm</b>	<b>RecNorm</b>
<b>86870</b>	49	8064000	79	0.0848	0.0004	0.0000
<b>95783</b>	89	19667000	84	0.1555	0.0010	0.0072
<b>918210</b>	70	24550500	81	0.1219	0.0012	0.0029
<b>86898</b>	88	13210020	81	0.1537	0.0007	0.0029
<b>532787</b>	72	38235520	88	0.1254	0.0019	0.0130
<b>86533</b>	153	49348900	82	0.2686	0.0025	0.0043
<b>910808</b>	132	46946500	88	0.2314	0.0024	0.0130

## LAMPIRAN C

### HASIL *CLUSTERING* PELANGGAN

**Tabel A 3** Cuplikan penentuan cluster pelanggan

Kode Langganan	Cluster
902347	1
9023594	1
96033	1
903953	1
515049	1
86848	1
730847	1
86626	1
96021	1
52835	2
86696	2
86611	2
86472	2
86878	2
929642	2
593708	2
579291	2
86621	2
9023593	2
96008	3
557173	3
920550	3
86789	3

Kode Langganan	Cluster
281820	6
95828	6
95781	6
95774	6
95853	6
95817	6
930414	6
95849	6
511120	6
926234	5
95799	5
927704	5
95820	5
641614	5
52493	5
95801	5
95842	5
95851	5
346471	5
95809	4
95812	4
95847	4
938504	4

629472	3
95955	3
86836	3
86690	3
9023592	3
96031	3
599074	4
118854	4
86732	4
902348	4
144515	4
86553	4
96029	4
902349	4
86613	4
86565	4
86911	5
831771	5
9023591	5
96012	5
720914	5
119056	5
902346	5
531569	5
927650	5
86709	5
614869	6
449826	6
639096	6

95775	4
95782	4
930012	4
95751	4
95818	4
95800	3
95796	3
86896	3
95803	3
565202	3
95805	3
95829	3
95840	3
95816	3
930813	2
95810	2
637240	2
9023567	2
9023568	2
9023571	2
9023569	2
9023570	2
281820	2
95828	2
95781	1
95774	1
95853	1
95817	1
930414	1

622352	6
86575	6
577589	6
733609	6
86902	6
504739	6
86895	6

95849	1
511120	1
926234	1
95799	1
927704	1
95820	1
641614	1

## BIODATA PENULIS



Penulis lahir di Surabaya pada tanggal 30 Agustus tahun 1993. Penulis telah menempuh beberapa pendidikan normal dari tingkat dasar sampai tingkat atas, yaitu SD Negeri Bebekan II, SMP Negeri 1 Surabaya, dan SMA Negeri 5 Surabaya.

Pasca kelulusan penulis pada tingkat atas di tahun 2011 dan meneruskan pendidikan ke jenjang perguruan tinggi di Jurusan Sistem Informasi

Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember (ITS) dan terdaftar sebagai mahasiswa dengan NRP 5211100102. Selama menjadi mahasiswa, penulis aktif dalam kegiatan kemahasiswaan dan aktif sebagai anggota Himpunan Mahasiswa Sistem Informasi, penulis juga aktif sebagai staff pada tahun kedua dan Ketua Departemen Hubungan Luar Badan Eksekutif Mahasiswa Fakultas Teknologi Informasi (BEM FTIf) pada tahun ketiga, dan disamping organisasi penulis juga sempat menjadi asisten pada mata kuliah Wawasan Teknologi (Wastek).

Pada jurusan Sistem Informasi, penulis tertarik dengan bidang pengolahan data dan data *warehousing*. Oleh karena itu penulis mengambil bidang minat Laboratorium Sistem Pendukung Keputusan dan Intelegensia Bisnis (Lab. SPK & IB) dengan topik Penggalian Data dalam pengerjaan tugas akhir. Penulis dapat dihubungi melalui email [akbarrv@gmail.com](mailto:akbarrv@gmail.com).